7 July 2006 | $10

Science

AAAS

Don Kennedy is the Editor-in-Chief of *Science*.

George M. Martin, M.D., is the Editor-in-Chief of SAGE KE. He is Professor of Pathology, Emeritus (active), and the Director Emeritus of the Alzheimer's Disease Research Center at the University of Washington in Seattle. He also serves as the Scientific Director of the American Federation for Aging Research. His lab investigates genetic modulations of aging and diseases of aging. E-mail: gmmartin@u.washington.edu

SAGE Lessons

The American Association for the Advancement of Science (AAAS) and *Science* ventured into new territory in 1999 with the launching of a Web-based "knowledge environment" or KE. KEs were envisaged as capturing major advances in rapidly developing and noteworthy areas of science—areas where progress is reported in widely scattered publications and thus is not readily accessible to researchers and educators in a comprehensive and integrated fashion. The first such initiative embraced the mushrooming field of cellular signal transduction research. STKE, the Signal Transduction Knowledge Environment (http://stke.sciencemag.org), has enjoyed a steady growth of subscribers and has become well entrenched within numerous scientific communities. Signaling pathways communicate with each other, and STKE has encouraged some healthy cross-talk among many of the disciplines in this field.

Two years after the emergence of STKE, Floyd Bloom, Ellis Rubinstein, and their colleagues at *Science* identified the biology of aging as a field of biomedical research with those same qualities of rapid development, high significance, and dispersed literature. Thus was born SAGE KE, the Science of Aging Knowledge Environment (http://sageke.sciencemag.org). The Ellison Medical Research Foundation provided an exceptionally generous initial grant. Other important contributions came from the Paul Glenn Foundation, the Merck Foundation, the Canadian Institute for Health Research, and the German Research Centre for Biotechnology. SAGE KE has seen subscriptions grow (its content has included freely accessible material as well as information relying on subscription support) as it forged a bridge between scientific disciplines, vigorously nurturing interactions and information exchange. It has been a bridge at many levels: providing timely articles about discoveries in numerous fields, publishing critical analyses of research, acting as a gateway to database information, serving educational tools, and offering a critical means for scientists to communicate with each other. And thanks to the hard work of Dan Perry and the Alliance for Aging Research, SAGE KE extended this bridge to policymakers grappling with hot-button aging issues through SAGECrossroads (www.sagecrossroads.net).

Unfortunately, there was insufficient support to sustain SAGE KE, and it will no longer post new content after June 2006. However, its wealth of material is archived and will remain accessible indefinitely (free to AAAS members) and searchable by PubMed. The history of SAGE KE repeats what is now a familiar story about the uncertain fates and longevities of new Web sites and electronic resources that depend on private and/or federal funding, and on the budgets of a research community that is already stretched to its limits. Survival depends on sustained funds, and granting agencies are more in the business of seeding rather than maintaining such projects. For those electronic resources providing freely accessible information, the struggle is especially tough. Recently, the Biomolecular Interaction Network Database—the world's largest free repository for proteomic data—lost its funding and curtailed its curation efforts. Its future remains unclear. Perhaps, like the Los Alamos preprint server—the favored repository of research by physicists, now located at Cornell University—it will find more secure support through other sources.

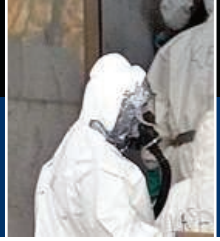
For those enterprises that offer unique combinations of information that is freely accessible, as well as content that is not funded by grants, endurance can be even trickier. Researchers and librarians have to make tough subscription decisions with increasingly strained budgets. Because electronic resources not only help individual research communities but can bridge them in new ways, they deserve support. The lesson for stakeholders among the scientific community, policymakers, and educators is that they need to take an active role in the viability of such enterprises. Otherwise, resources they consider valuable may simply become electronic history.

As for SAGE KE, we are exceedingly grateful to the communities that contributed and for the exceptionally talented editors and writers that joined it. On the editorial and journalistic side, Kelly LaMarco, Evi Strauss, Heather McDonald, John Davenport, and Mitch Leslie have our lasting thanks, as do our numerous scientific advisors and contributors, including many postdoctoral fellows who reported from the trenches. SAGE KE may have taken a break from the business of further development. But perhaps, as biogerontologists would say, this is only a period of pause—a "dauer" phase—in its life history.

–George M. Martin and Don Kennedy

10.1126/science.1131700

Published by AAAS



Gaps in BioShield
28

Toward harmony in the Himalayas
30

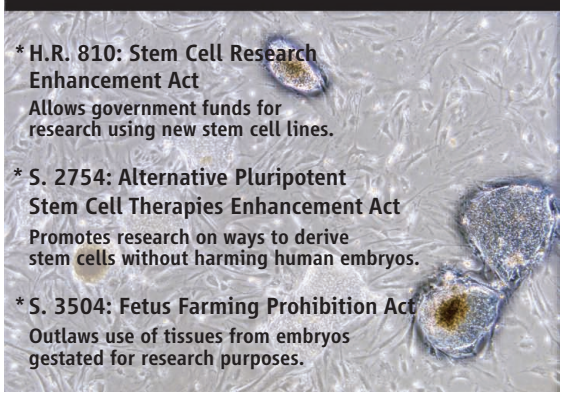
STEM CELL RESEARCH

Senate Prepares to Vote at Last On a Trio of Stem Cell Bills

Senate leaders have formally agreed to allow a vote—possibly this month—on a bill that would allow federally funded researchers to work on newly derived lines of human embryonic stem (ES) cells. The bipartisan deal announced last week was painstakingly cobbled together over the past few months to placate opponents by including one bill that would promote "alternatives" to embryo destruction for obtaining stem cells and another that would outlaw "embryo farms."

Supporters of stem cell research have lobbied hard for an up-or-down vote in the Senate on a bill, passed in May 2005 by the House (H.R. 810), that would allow federally funded researchers access to cell lines derived after the presidentially imposed cutoff date of 9 August 2001 (*Science*, 3 June 2005, p. 1388). Last summer, Senate Majority Leader Bill Frist (R–TN) reversed his previous opposition to human ES cell research and said he supported H.R. 810. But as the months rolled by without a Senate vote, many stem cell boosters began to worry that Frist, a physician who is leaving the Senate at the end of the year for what is expected to be a run for president, might be dodging the issue for political gain.

The Senate is expected to vote this summer on these three bills:



* **H.R. 810: Stem Cell Research Enhancement Act**
  Allows government funds for research using new stem cell lines.
* **S. 2754: Alternative Pluripotent Stem Cell Therapies Enhancement Act**
  Promotes research on ways to derive stem cells without harming human embryos.
* **S. 3504: Fetus Farming Prohibition Act**
  Outlaws use of tissues from embryos gestated for research purposes.

But it turns out that Frist, along with stem cell advocates Senators Arlen Specter (R–PA) and Tom Harkin (D–IA), has been working hard to win a so-called unanimous consent agreement that commits members to the terms of the vote. The last piece to fall into place, apparently, was convincing fellow physician Tom Coburn (R–OK) to drop his own "alternatives" bill in favor of the agreed-upon legislative troika. Frist's office says he intends to schedule a vote before the Senate goes into its August recess.

Under the agreement, H.R. 810 will be buffered by two bills designed to appeal to opponents of embryo destruction. One (S. 2754), co-sponsored by Specter and Rick Santorum (R–PA), calls on the National Institutes of Health (NIH) to promote research on finding ways to derive pluripotent cells other than from embryos. The bill would only reinforce current NIH policies, NIH stem cell czar James Battey told senators last week at a hearing on the legislation. The other measure (S.3504), co-sponsored by Santorum and Sam Brownback (R–KS), prohibits trading in tissues from human fetuses "gestated [in humans or animals] for research purposes." This is already prohibited under federal funding rulesnd wo auld in any case be ethically taboo for legitimate researchers. Because the bills are not mutually exclusive, the Senate could easily pass all three. ▶

International Standards Proposed for Stem Cell Work

TORONTO—Scientists who work on stem cells have proposed draft guidelines to set ethical standards for researchers around the world. The guidelines, which are the work of an international committee, lay out ground rules for work with embryos and the cells derived from them. The document also recommends ethical standards for obtaining sperm, eggs, embryos, or other cells from human donors.

The guidelines are consistent with those set out by the U.S. National Academies last year (*Science*, 29 April 2005, p. 611), but "we extend and refine those principles" for the international community, George Daley of Harvard Medical School in Boston said at a meeting* here last week. Daley, who headed the drafting committee with 30 members from 14 countries, says the document should ease collaborations between scientists who live in regions of the world with different laws and local regulations regarding use of embryos or informed consent of tissue donors.

The guidelines recommend that certain types of research, such as derivation of new embryonic stem cell lines or generation of chimeric animals, be subject to special review by an independent panel. In some cases, the panel may be at the investigator's institution; others might be governed by a regional or national review. The guidelines also set standards for sharing research materials including reagents, animal strains, and cell lines and urge scientists to deposit new cell lines at national or international cell banks.

Committee members said their most intense debates concerned how to fairly compensate women who donate oocytes for research. People who donate bone marrow for research, for example, are usually paid for their time, discomfort, and inconvenience, but several committee members felt strongly that oocyte donors should not be offered any compensation beyond reimbursement for their expenses. In the end, the committee agreed that local review boards should ensure that compensation does not "constitute an undue inducement" but otherwise left the final decision up to local laws and practices.

The committee plans to draw up template documents for material transfer agreements and informed consent for the donation of cells or embryos. Daley says such templates would have made the work dramatically easier as he and his colleagues prepared to begin human nuclear transfer experiments this spring.

ISSCR members have 60 days to comment on the draft, which has been posted on the society's Web site. The committee hopes to issue a final document by the end of the year. **–GRETCHEN VOGEL**

* International Society for Stem Cell Research 4th Annual Meeting, 29 June–1 July, Toronto, Canada.

CREDIT: UNIVERSITY OF WISCONSIN, MADISON

Published by AAAS

From observer to advocate 32

Chemicals that bug bugs 36

Remembering Hans Bethe 39

The agreement has successfully divorced the matter of generating new cell lines (from excess embryos in fertility clinics) from an issue with which it has often been conflated: generation of cell lines through research cloning (otherwise known as somatic cell nuclear transfer). Earlier scenarios of how H.R. 810 might be brought before the Senate included a Brownback-sponsored bill that would outlaw all forms of cloning. Research supporters feared that President George W. Bush would veto H.R. 810 and sign the anti-cloning bill into law, leaving them worse off than under present circumstances.

There are now 21 human ES cell lines available to federally funded researchers. But scientists want more: Cell lines get corrupted over time by genetic mutations; the available ones were all cultivated using animal feeder cells, which limits potential use for humans; and researchers want to be able to work with lines containing genes for specific diseases.

Representative Mike Castle (R–DE), a prime mover in getting H.R. 810 through the House, predicts a "momentous" debate and praises Frist for his "thoughtfulness and commitment." Adds Kevin Wilson of the American Society for Cell Biology in Bethesda, Maryland, "I was afraid he wasn't going to be able to get it through. The one thing we knew is it had to be a clean bill."

Each bill needs a filibuster-proof 60 votes to win passage. But Wilson notes that retaining the exact language that the House passed was also vital because any changes would force the bill into a conference with the House, where it could be swamped with amendments or delayed indefinitely.

The Senate's willingness to take up H.R. 810, he and others note, also may reflect polls showing that the vast majority of the public supports it. Still, that may not be enough. White House spokesperson Blair Jones says Bush won't budge on the issue. "He does not believe we are forced to choose between science and ethics," says Jones. "This crosses an important moral line."

–CONSTANCE HOLDEN

STEM CELL RESEARCH

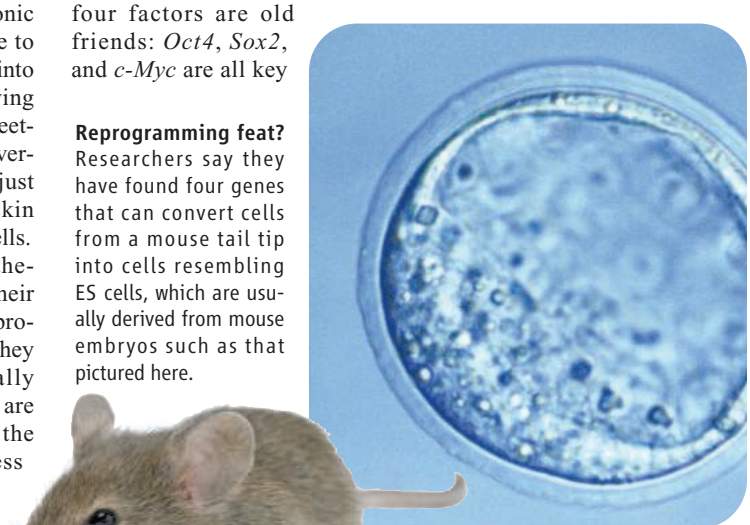
Four Genes Confer Embryonic Potential

TORONTO—One of the biggest questions in stem cell biology is how the cloning process manages to turn back the clock of differentiated cells, resetting them to their embryonic potential. Ideally, researchers would like to find a way to convert adult cells directly into embryonic stem (ES) cells—without having to create an embryo at all. At the ISSCR meeting* here, Shinya Yamanaka of Kyoto University in Japan reported that upregulating just four genes can apparently turn mouse skin cells into cells that closely resemble ES cells.

Yamanaka and his colleagues hypothesized that the factors that give ES cells their unique properties might also be able to reprogram adult cells to behave like ES cells. They identified 24 genes that are specifically expressed in mouse ES cells. Six of them are well known to ES cell researchers, but the team also fingered 18 genes that are less famous in the field but are consistently turned on in ES cells.

The team used viral vectors to introduce extra copies of the 24 genes into skin cells taken from mouse tail tips. When they inserted extra copies of all 24 genes, they found that a small percentage of cells that took up the genes did indeed seem to take on characteristics of ES cells. But no single gene introduced alone was able to manage the transformation.

Through a process of elimination, the team whittled down the candidates to a suite of just four genes that, when introduced together into the tail-tip cells, could produce colonies of ES-like cells. As Yamanaka described, three of the four factors are old friends: *Oct4*, *Sox2*, and *c-Myc* are all key genes in both early embryos and ES cells. The fourth is one of the 18 lesser known genes the team had identified. Yamanaka did not name it but said it is a transcription factor that until now has not been recognized as playing a major role in ES cells.



Reprogramming feat? Researchers say they have found four genes that can convert cells from a mouse tail tip into cells resembling ES cells, which are usually derived from mouse embryos such as that pictured here.

The ES-like cells the group produced with the four introduced genes seemed to have almost all the key properties of ES cells derived from embryos. They formed several kinds of tissue in the culture dish, formed tumors called teratomas when they were injected under the skin of immune-compromised mice, and seemed to contribute to almost all tissues when they were mixed with mouse embryos and allowed to develop—all classic characteristics of ES cells.

Yamanaka says his group has not yet tried the technique with human cells. Because of differences in human and mouse embryo development, he says, it's possible that a different set of genes would be required to reprogram human cells.

Other researchers at the meeting were impressed. "It's huge," says Kevin Eggan of Harvard University, who also works on reprogramming. He notes that the process is not yet very efficient; the four introduced genes managed to reprogram just 1 out of 1000 cells that received them. That suggests that the four genes are perhaps not the whole story, and that another factor could improve the efficiency of the process. "But this is the litmus test" for finding the genes that are essential for reprogramming, he says.

–GRETCHEN VOGEL

*International Society for Stem Cell Research 4th Annual Meeting, 29 June–1 July, Toronto, Canada.

CREDITS: (TOP TO BOTTOM): CLOUDS HILL IMAGING LTD./CORBIS; DIGITAL ZOO/GETTY

Published by AAAS

BIOTERRORISM

BioShield Is Slow to Build U.S. Defenses Against Bioweapons

Developing vaccines against potential bioweapons such as smallpox and Marburg virus is tough going for small companies. But it's even harder when their comrade-in-arms on the front lines, a $5.6 billion federal program called BioShield, is AWOL.

AlphaVax, a North Carolina biotech, has already received $26 million from the National Institutes of Health (NIH) to explore how its technology, which uses a genetically modified alphavirus to stimulate a broad immune cell response against chosen microbes, can be turned into vaccines against various biothreats. But if an initial clinical trial of its prototype botulinum toxin vaccine goes well, it will likely need $100 million or more for larger clinical trials and scale-up—money that the 70-person company doesn't have on hand.

Enter BioShield. When President George W. Bush proposed the procurement program in his 2003 State of the Union address, he reasoned that the promise of lucrative sales to the government would allow companies to keep their scientists employed and their manufacturing plants rolling. But things haven't worked that way. Companies often need more than NIH funding to be ready to bid for a BioShield contract, and even companies that succeed are paid only after they deliver a vaccine or therapy.

Those twin "valleys of death" aren't the only flaws in BioShield, which critics say is understaffed and badly managed. The six contracts awarded to date, for $1.7 billion, are aimed at countermeasures against just three threats, and in most cases the amounts are modest (see table, above). Both houses of Congress are weighing legislation aimed at correcting BioShield's apparent flaws. But those potential remedies offer scant immediate hope to AlphaVax CEO Peter Young. "We don't know if there's a next step" for the 8-year-old company's biodefense products, he frets.

BioShield Contracts Awarded

Product	Company	Amount (millions)	Doses (millions)
Pending FDA approval			
rPA anthrax vaccine	VaxGen Inc.	$878	75
Botulism antitoxin	Cangene Corp.	$363	0.2
ABthrax anthrax therapeutic	Human Genome Sciences	$165	0.02
FDA approved			
Liquid potassium iodide*	Fleming & Co.	$16	4.8
AVA anthrax vaccine	BioPort Corp.	$243	10
Ca-DTPA, Zn-DTPA chelators*	Akorn Inc.	$22	0.45
Total		**$1686**	

*Radiological/nuclear attack.

Thin shield? BioShield has spent more than half of its money to date on a single contract for an rPA anthrax vaccine.

Inside the black box

Drug development is a dodgy proposition at best. But developing biodefense countermeasures is even more of a crapshoot because there's no certainty they'll ever be used. Having the government as a customer is essential, say companies. The 10-year Project BioShield created by Congress in 2004 is meant to build on the now-$1.7-billion-a-year investment in basic biodefense research by NIH's National Institute of Allergy and Infectious Diseases (NIAID).

Getting a BioShield contract isn't simple. The bioagent for which a company is seeking a countermeasure must be on the Department of Homeland Security's list of public health "material threats." The company must promise that the product will be available "in sufficient quantities" within 8 years. And the firm doesn't get paid until the product is delivered.

Even before Congress created BioShield in 2004, experts warned that big drug companies would turn up their noses because the potential profits would be minuscule compared to those from, say, a blockbuster drug for lowering cholesterol. And indeed, small biotech companies have been much more likely to chase after BioShield contracts.

But biotechs often struggle to attract the investor funding needed to get a product ready for BioShield. NIH funding rarely covers all the preclinical studies necessary to qualify, says Robert Housman, a homeland security consultant in Washington, D.C. And scaling up manufacturing and conducting later-stage clinical studies of a typical drug or vaccine costs "at least $100 [million] to $150 million," he says. NIAID has helped fund those steps for two products: new smallpox and anthrax vaccines. But an institute official says that regularly paying for development could devastate NIAID's budget.

Those practices have put some companies in a holding pattern. "We have been frustrated by the limitations of the current system," Bruce Cohen, CEO of Cellerant Therapeutics Inc., told a House panel in April. Cellerant has used NIAID funds to collect promising animal data on a cell-based treatment for restoring the immune systems of victims of radiological attack, he says, but the company needs another $10 million to $20 million for scale-up and clinical trials.

Even if a company manages to cover development costs, a sale to the Department of Health and Human Services (HHS), which awards BioShield funds, is hardly a shoo-in. "The system is not transparent, and the government doesn't tell you what it wants," says Francesca Cook, vice president of policy and government affairs at PharmAthene, an Annapolis, Maryland–based biotech that has been waiting a year for HHS to request new proposals for antianthrax antibodies. In contrast, Cook says, the Defense Department has been much more open about its interest in the company's nerve agent treatment.

Critics lay part of the blame on HHS's Office of Public Health Emergency Preparedness, which manages BioShield. Tara O'Toole, director of the University of Pittsburgh Center for Biosecurity in Pennsylvania, estimates that the 50-person office needs to triple its size and employ more experienced hands. Many potential bioweapons, such as Marburg virus, are missing from BioShield's current eligibility list, and it's not clear if or when they will be added.

BioShield has struggled to manage existing contracts, too. The most notorious case involves VaxGen, which in 2004 received an $878 million contract to supply recombinant anthrax vaccine. The company had to reformulate its vaccine after learning that its active ingredient had a short shelf life. But company officials say the government moved the goalposts this spring by asking for additional clinical and animal tests and insisting that VaxGen should pay for the extra work, which has pushed back the 2006 delivery date by 2 years.

"Contemplating that interaction doesn't whet industry's appetite" for working with the government, says AlphaVax's Young. Neither does the fact that HHS may order less of a product than a company expects. For instance, last month HHS contracted with Human Genome Sciences Inc. in Rockville, Maryland, for just 20,000 doses of an anthrax treatment rather than the possible 100,000 mentioned in an earlier agreement.

Corrective action

Congress has begun to address some of these concerns. In December, legislators agreed to insure companies against liability for vaccines, a major omission in the original legislation.

Building on a measure passed last year by the Senate Health, Education, Labor, and Pen-

CREDITS: (TABLE) HHS OFFICE OF PUBLIC HEALTH EMERGENCY PREPAREDNESS; (PHOTO) RANDY DAVEY/REUTERS

Published by AAAS

Biopreparation. BioShield is supposed to help U.S. officials respond better to incidents such as the anthrax-laced letters mailed to legislators in 2001.

sions Committee, Senator Richard Burr (R–NC) has called for a new Biomedical Advanced Research and Development Agency (BARDA) that would serve as a "single point of authority" within HHS for developing countermeasures for biodefense and natural pandemics. BARDA would bridge the valley of death by spending up to $500 million of BioShield for "advanced development." Burr's bill, S. 2564, would also allow BioShield payments for manufacturing and for achieving certain milestones, permit the HHS secretary to hire contractors at salaries exceeding federal levels, and set up an advisory board, with members from industry, to identify new biological threats. The bill also tones down controversial exemptions to open-records laws in the previous version, although some of these provisions remain.

A companion bill in the House, introduced by Representative Mike Rogers (R–MI) last month, would provide $1.07 billion over 3 years for BARDA. A Senate staffer says that the money would come from shifting $160 million now tagged for advanced bioweapon product development in NIH's 2007 budget and other funds already appropriated to combat pandemic flu. The staffer says the Bush Administration requested the reallocation.

HHS officials say they are "quite supportive of the intention" of both the Burr and Rogers bills and are already working on a strategic plan that the legislators have proposed. But observers say a bigger pot is needed, too. The Alliance for Biosecurity, a collaboration of companies and O'Toole's center, cites a 2000 Defense Science Board study that identifies at least 19 major biothreats and notes that developing a drug or vaccine costs $800 million. "The government has not recognized the reality and the scale of the biodefense threat," says O'Toole. Still, she calls Burr's bill "a milestone in this evolving and emerging discussion." **–JOCELYN KAISER**

SCIENTIFIC PUBLISHING

A Mixed Bag of U.K. Open-Access Plans

CAMBRIDGE, U.K.—The open-access movement chalked up a victory in Britain last week, but it did not get the universal mandate for free release of research papers that some advocates want. In a long-delayed policy statement on 28 June, the executive board of Research Councils U.K. (RCUK), an umbrella organization for government funding bodies, said that all peer-reviewed journal papers produced by publicly funded research must be made available for free soon after they're completed. Exactly what that means was not specified, and RCUK left each research council to set its own rules. In coordinated announcements, some set out hard-edged policies whereas others said they were still debating what to do.

The most stringent policy came from the Medical Research Council (MRC), source of roughly $400 million in annual biomedical grants. It declared that any papers accepted by a peer-reviewed journal must be deposited "at the earliest opportunity—certainly within 6 months—in PubMed Central," the free public archive run by the U.S. National Library of Medicine. A mirror U.K. archive is under construction, MRC says, and when it is ready all MRC-funded papers will go there. The rule applies to grants awarded from October 2006 onward, and an MRC spokesperson explains that the 6-month clock begins to run from a paper's publication date.

The Biotechnology and Biological Sciences Research Council, which backs more applied projects, took a softer line. It stipulated that papers should be submitted only "at the earliest opportunity" and placed in "an appropriate e-print repository." Other statements were even less precise. The Engineering and Physical Sciences Research Council, for example, said that "the issues are complex" and must be examined further; the council will await the results of a study due in 2008. The Particle Physics and Astronomy Research Council said merely that it "will consider … what changes might now need to be made."

The news that MRC is setting hard deadlines won praise from some advocates of open access, such as BioMed Central, a London-based commercial scientific publisher funded by billing authors rather than readers. "It was a very important step," says publisher Matthew Cockerill. "Not many other funding ▶

CREDIT: WIN MCNAMEE/REUTERS

Panel: Extensive Sudbø Fraud

Most of the published papers by oral cancer expert Jon Sudbø of the University of Oslo's Norwegian Radium Hospital are bogus, according to an investigative panel. "The bulk of Jon Sudbø's scientific publications are invalid due to fabrication and manipulation of raw data," says the 30 June report by an independent group headed by epidemiologist Anders Ekbom of the Karolinska Institute in Stockholm.

This winter, Sudbø confessed that he had invented data in a 2005 *Lancet* paper on detecting oral cancer, leading to a retraction of that paper and casting suspicion on other high-profile papers (*Science*, 27 January, p. 448). The Ekbom panel found, for example, that nine of 150 patients in one study were fictitious and that another 69 "should have been excluded" because they had already been diagnosed with cancer. The report says Sudbø's doctoral dissertation and papers based on its data should be retracted.

Stein Evensen, dean of the University of Oslo medical faculty, called the investigation "very accurate" and pointed to a silver lining: None of Sudbø's 60 co-authors is implicated in his misdeeds. But the Ekbom panel found an apparent "systemic failure" by the Oslo hospital to stem the fraud. Evensen will propose that the university withdraw Sudbø's doctorate and adopt better oversight procedures. Neither Sudbø nor his attorney could be reached for comment.

–ELIOT MARSHALL

To the Moon, Barney

NASA should stick to its exploration plans, House lawmakers declared last week. When the agency's $16.7 billion budget for 2007 went to the House floor, Representative Barney Frank (D–MA) urged colleagues to block $700 million from use for human Mars exploration, calling the spending "an extravagance" and "a psychological stunt" with no scientific value. But Frank's proposal lost, 259–163. The Senate is preparing its own version of NASA's budget, which likely will increase funding for science, aeronautics programs, and pork-barrel projects. NASA backers are hoping that a safe space-shuttle mission this week will prove that the agency still has the right stuff for human exploration.

Meanwhile, after a scare last week, the Hubble Space Telescope's main camera survived the failure of a power source aboard the orbiting spacecraft. Engineers used an alternative power supply to bring the Advanced Camera for Surveys back into research operation.

–ANDREW LAWLER

Published by AAAS

institutions have gone as far." MRC is following the lead of the largest U.K. research charity, the Wellcome Trust, which spends $890 million per year on biomedicine. Last year, Wellcome told all grantees they must submit accepted research papers to PubMed Central so that they could be released "no later than" 6 months after official publication. According to Cockerill, MRC's decision means that nearly all new biomedical papers in Britain will come under open-access rules. In the United States, meanwhile, Congress is considering a bill that would require papers funded by the National Institutes of Health be placed in PubMed Central within 12 months of publication (*Science*, 16 June, p. 1585).

Cockerill was also encouraged by a sentence in the MRC announcement indicating that the government may pay the costs of publishing in open-access journals. (The Wellcome Trust already reimburses authors.) But others worry that the costs of moving to open access are growing. BioMed Central charges between $1380 and $1750 per article; Cockerill reports that at these rates, it is "almost breaking even." The U.S.-based *Public Library of Science*, which is subsidized by foundation grants, recently raised its maximum fee from $1500 to $2500 per article. And an experimental open-access journal launched in June by the Royal Society in London, *EXiS Open Choice*, is planning to charge $415 to $553 per printed page. No-subscription publishing can be very expensive, says London-based consultant Mary Waltham, "and many publishers were disappointed" last week to see that RCUK did not back its open-access policy with a promise of new financial support.

RCUK recognizes that it is moving into unexplored territory. To investigate the costs and risks, it is recruiting three major companies (Macmillan, Elsevier, and Blackwell) to join in a 2-year analysis of the impact of mandating free release of journal articles. It will review these policies in 2008.

–ELIOT MARSHALL

SCIENCE FOR PEACE

Across a Political Divide, Researchers Converge on Himalayan Plan

DEHRA DUN, INDIA—A few dozen geoscientists met here in the foothills of the Himalayas last week to lay the groundwork for a bold initiative that would bring researchers from India and Pakistan together on joint projects in Kashmir. But there was a glaring hitch: Their Pakistani colleagues were on the other side of the border.

It wasn't supposed to be like this. After months of delicate planning, scientists had been set to gather in Islamabad at the end of May to hammer out a research plan for the western Himalayas, in particular the Karakoram Mountains. "There can't be a better natural earth science laboratory than the high Himalayas," says John "Jack" H. Shroder, a geoscientist at the University of Nebraska, Omaha, and co-organizer of the meeting, funded in part by the U.S. National Science Foundation (NSF). A centerpiece was to have been a discussion of a "science peace park" centered on the Siachen Glacier, a high-altitude graveyard for troops on the disputed border.

At the last minute, however, the Pakistan government withdrew its support for the meeting, citing security concerns (Science, 26 May, p. 1117). The cancellation appeared to be collateral damage from the glacial pace of India-Pakistan talks on Siachen demilitarization, says Harry Barnes, a former U.S. ambassador to India who is advising NSF on the initiative.

Event organizers regrouped as best they could. On 31 May, 35 Pakistani and six U.S. scientists met in Islamabad to cobble together a research manifesto for the western Himalayas. A similar wish list was produced in Dehra Dun by a few dozen Indian scientists and colleagues from Canada and the United States. Neither meeting had local government support; Pakistan denied visas to Indian scientists, whereas Dehra Dun organizers say that time was too short after the May debacle to seek visas for Pakistani counterparts.

A top priority of all sides is to better understand Himalayan geodynamics. Accentuating the need for such studies is the earthquake that struck Kashmir on 8 October 2005, killing more than 100,000 people. The magnitude-7.6 quake "was a wake-up call … that temblors do not respect national boundaries," says Shroder. One nasty surprise was that the quake's epicenter—the Muzaffarabad fault—was not known to be active, he says.

Researchers called for the installation of a seismic network to better map tectonic activity in the western Himalayas. That would require unprecedented cooperation between Indian and Pakistani security forces, says Michael P. Bishop, a geoscientist at the University of Nebraska, Omaha. Researchers also hope to undertake active seismic profiling, in which explosives are detonated in deep holes. The vibrations reveal rock composition and fault structure—vital to refining maps of seismic risk. Mary Leech, a geologist at San Francisco State University in California, has tried to launch such work with Indian colleagues. "We have been stopped because of the complex political problems," she says. "Carrying out even small explosions in border areas can be very problematic."

Both meetings generated other ideas under the banner of a "Siachen Science Laboratory." Proposals include probing the potential effects of climate change on monsoons and documenting the retreat of border

CREDIT: CHANNI ANAND/AP

A world apart. Tensions between India and Pakistan are hampering plans for geophysics research in the western Himalayas, including the iconic Siachen glacier.

glaciers. Barring a rapprochement on the Siachen military issue, the scientists intend to meet next year in a neutral venue—possibly in the Italian Alps—to merge research agendas and seek sponsors.

Although the absence of Pakistani researchers was acutely felt in Dehra Dun, Shroder urged scientists to keep their spirits up. "Just keep pushing the edges, and little by little, good science can be done," he said. And as Baldev R. Arora, director of the Wadia Institute of Himalayan Geology in Dehra Dun, optimistically predicted, "Opportunities for collaborative work among all Himalayan neighbors can only increase." It may take a significant thaw between cold warriors for those hopes to become reality.

–PALLAVA BAGLA

JAPAN

Top Chemist Accused of Funds Misuse

TOKYO—One of Japan's most prominent scientists is facing allegations of misappropriating funds. Kazuko Matsumoto, who was poised to become head of an international scientific society, has acknowledged improperly handling payment requests but claims the money was used for research purposes. The case has prompted government officials to call for stricter oversight of research grants.

Probe target. Kazuko Matsumoto.

Tipped off by a whistleblower, Waseda University in Tokyo last April launched an investigation into the financial dealings of Matsumoto, 56, a chemist in the university's School of Science and Engineering. In a 5-page interim report released on 23 June, the investigating committee said that between 1999 and 2003, Matsumoto drew $128,000 from publicly funded research grants to pay wages of six students she claimed were working part-time in her lab. According to the report, the students had not worked for Matsumoto, and the money was channeled into a personal bank account. She later transferred $78,000 into a personal investment fund managed by a stock brokerage and spent another $10,000, the report states.

The report says Matsumoto admitted improperly handling the student payment requests but said the money was used for student travel expenses and to purchase supplies. It notes that she has agreed to return the money. The report also alleges that Matsumoto may have made more than $200,000 worth of fictitious purchases from a biotechnology firm she was working for as a part-time director, but the committee acknowledges it lacks conclusive evidence and that Matsumoto has denied any wrongdoing. The probe is continuing.

Matsumoto declined to comment. "When the time comes, I might have something to say, but right now I am not saying anything to the media," she told *Science*.

The allegations have dealt a severe blow to Matsumoto's career. A few days after the report's release, she resigned as vice-president of the International Union of Pure and Applied Chemistry (IUPAC). She was due to ascend to the presidency in January 2008, when she would have been the first female president in the society's 87-year history. "We're sorry this happened to her," says John Jost, IUPAC executive director. The society's governing body will decide what to do about the vacant post when it meets this fall.

From 2002 until last January, Matsumoto was a member of the Council for Science and Technology Policy, Japan's highest science advisory group. She had also served on a Ministry of Education committee on research misconduct.

The committee's report leaves a number of questions unresolved. It notes, for example, that a separate Waseda panel is investigating allegations that data were fabricated in a paper that Matsumoto wrote with a collaborator. The report also declares that officials in the School of Science and Engineering had learned of concerns about Matsumoto's handling of funds 2 years ago but hadn't followed up. Waseda President Katsuhiko Shirai is heading a new panel to create procedures for preventing the misuse of research funds.

Government officials have been roused to action. Referring to the Matsumoto case on 29 June, Iwao Matsuda, minister for Science and Technology Policy, promised to have guidelines on preventing abuse of research funds drawn up this summer. If organizations don't fall in line, he stated, "we will withhold their competitive grants and take other measures."

–DENNIS NORMILE

CREDIT: IUPAC

Enviro Journal Staying Put

National Institute of Environmental Health Sciences (NIEHS) Director David Schwartz has decided not to privatize the field's most important journal. Schwartz said last fall that it's unusual for a federal agency to publish a major journal and that privatizing *Environmental Health Perspectives* would give it greater independence (*Science*, 2 December 2005, p. 1407). But hundreds wrote in favor of keeping the journal at NIEHS. They warned that a private publisher might influence its content.

"We were persuaded," says Schwartz. But changes he's planned include appointing an outside scientist as editor-in-chief and halving the news section. The decision overall is "good news," says environmental health researcher David Ozonoff of Boston University.

–JOCELYN KAISER

Warning on Wave Warnings

NEW DELHI—Last week, UNESCO declared a long-awaited Indian Ocean Tsunami Warning System to be "up and running," pointing at 25 new seismic stations, three deep-ocean buoys, and 24 national information centers for distributing advisories. But experts say that data-integration problems persist and large gaps in coverage are yet to be filled, especially the South China Sea and the Makran subduction zone south of Pakistan, with India's independent system incomplete for at least 15 months. Thailand plans to deploy its first deep-ocean buoy in December, and a German-Indonesian sensor system remains a work in progress. "We have a long way to go," says Costas Synolakis, director of the Tsunami Research Center at the University of Southern California in Los Angeles.

–PALLAVA BAGLA

Cloning Proposed for Japan

TOKYO—A panel has recommended that Japan lift its ban on therapeutic cloning by allowing researchers to use surplus eggs from fertility treatments to obtain embryonic stem cells. Scientists welcomed the proposed change, contained in an interim report last month to the Ministry of Education. "The importance of this technology for the study of human embryology, human oncology, and drug discovery will increase," says Shin-Ichi Nishikawa, a stem cell researcher at the RIKEN Center for Developmental Biology in Kobe.

The new rules, now open for public comment, will be vetted by the prime minister's Council for Science and Technology Policy.

–DENNIS NORMILE

Published by AAAS

Simron Singh had earned a reputation as a top expert on the Nicobarese. Then disaster struck, and Singh made a fateful decision: to ditch any pretense of objectivity and help rebuild their culture and their lives

After the Tsunami: A Scientist's Dilemma

BANGKOK—Two days after a massive tsunami pummeled southern Asia in December 2004, a message picked up by shipboard radio reached Simron Jit Singh at his parents' home in Lucknow, India. It was from Rasheed Yusuf, a close friend in the Nicobar Islands, a little-known archipelago a few hundred kilometers from the earthquake that triggered the tsunami. The news was bad: "Central Nicobars entirely washed out. … Do something as soon as possible."

Singh, a human ecologist and anthropologist at the Institute of Social Ecology in Vienna, had spent the previous 5 years chronicling the indigenous Nicobarese. He lived among them for weeks at a stretch, earning their trust and gathering a wealth of information. Now the society itself seemed to be slipping away: Out of a population of 30,000, about 4500 had perished in the tsunami; another 5000 were missing and presumed dead. Nine of every 10 homes on the 24-island chain were reduced to splinters. The islanders' economic lifeblood, coconut palms, was virtually wiped out. Most insidious, nearly every artifact—irreplaceable ossuaries and other relics preserved for generations—had been washed away.

The tsunami left the numbed survivors at a crossroads. Leaders were torn between either trying to restore their cultural identity or accelerating a fitful integration with the outside world in which many Nicobarese had already adopted Western clothing and other trappings of modern life, from television to pop music. Tribal elders sought the counsel of an outsider they knew they could trust: Singh.

That left Singh facing his own moment of truth. Until then, he had remained loyal to the scientific creed of minimal intervention. Yes, many research subjects had become friends and confidants. And yes, his work was influencing their lives in subtle ways. Now, however, the Nicobarese were asking Singh for much more: to cross the line between observer and participant and help make decisions that could determine whether the islanders would retain centuries-old traditions as a facet of their rapidly changing lifestyle.

It didn't take Singh long to decide. He flew to the Nicobars in late January 2005 and, since then, has assisted the islanders in restoring their culture and reshaping their economy. "He has literally single-handedly brought to the world's attention the cultural, social, and economic plight of the Nicobarese," says Mahendra Shah, a sustainable-development expert at the International Institute for Applied Systems Analysis (IIASA) in Vienna. Shah and others applaud the path Singh chose. "He could not in good conscience do anything else," says Pernille Gooch, a human ecologist at Lund University in Sweden.

Singh had looked into his heart and knew, he says, that "I really had no choice."

Listening and learning

Singh, 36, became involved with the Nicobars by chance. His first project as a student at Lund in the mid-1990s had been a study of the Van Gujjars, a tribe of nomadic buffalo herders in the Himalayas. Then one day in 1998, an Indian historian tracked him down at a Van Gujjars camp and urged him to study the Nicobarese. He was intrigued.

Working in this remote community, Singh learned, would not be easy. Many Nicobarese view outsiders with suspicion—for good reason. They have been host to a series of unwanted visitors. Situated on the trade routes between India and East Asia, the archipelago was colonized by Denmark in 1756, then by Austria, and finally by Great Britain, which held the Nicobars until India's independence in 1947; they are now formally part of India. But it remained a "marginalized society that few people had heard about," says Gooch.

To protect the indigenous peoples of the Nicobar Islands, as well as those on the Andamans to the north, India places strict controls on outsiders' access. Singh has Indian nationality, which helped him get a research permit, but he had to promise not to divulge any information deemed sensitive to Indian security. Yet his nationality was also a liability: The Nicobarese are wary of Indian traders on the islands. What's more, tribal elders held scien-

CREDIT: SIMRON SINGH

Published by AAAS

Culture shock. A young man tests his mettle against a tethered pig in a festival on Chowra Island in 2001 (*left*). The same spot, viewed from a different angle, is shown here after the 2004 tsunami.

tists in low esteem. When Singh visited in April 1999, the first impression of Ayesha Majid, chief of Nancowry Island, was "that he was just like other people who come, hear, write, and leave." Singh spent 2 months on nearby Trinket Island and promised to come back. "We were rather sure that he wouldn't," Majid says.

Singh's return dumbfounded the Nicobarese. "After nearly a year, we see Simron walking towards my home with a black bag on his back and a smile on his face," Majid says. "His respectful behavior touched us all." She says he would partake in raw fish and pork with the islanders, observe ceremonies even in the dead of night, and quiz them "endlessly" about their culture and traditions.

Singh kept returning, season after season, fascinated by a "very rich" culture preserved by limited contact with outsiders on some of the islands. He witnessed unique traditions, such as the annual pig festival, *Panuohonot*, which features a coming-of-age rite in which young men prove their valor in hand-to-hoof combat with pigs, and an ossuary celebration, *Kinruaka*, in which ancestral bones are dug up and reburied. Singh learned how heavily the Nicobarese depended on coconuts. A third of production was reserved for raising pigs, which grew extra-fat on the oil-rich diet. "To be a Nicobarese means to have pigs. The more pigs you have, the better off you are," Yusuf says.

For his thesis at Lund University, Singh explored the social metabolism of Trinket, unraveling the island's life in a monograph, *In the Sea of Influence*. "It's a thorough political, economic, and environmental history of these forgotten islands," says Joan Martínez-Alier, president of the International Society for Ecological Economics and a professor at the Universitat Autònoma de Barcelona in Spain.

The more Singh got to know the islanders, however, the more he worried about their future. Long before the tsunami, he decided they needed better links to the outside. Tensions between Nicobarese and ethnic Indians over land use and trade were likely to worsen over time, he felt. And the one-dimensional coconut economy made the Nicobarese "very, very vulnerable," says Yusuf. "We needed an alternative economy."

Singh concluded that the more information the Nicobarese had, the better armed they would be to make decisions about their future. He helped arrange invitations from Vienna and Lund for Yusuf to visit Europe. The trips were eye-opening, says Yusuf, who was one of the first Nicobarese to visit a foreign land and now handles external affairs for the tribe.

Privately, Singh had reservations about his evolving relationship with the Nicobarese. With his supervisor, Marina Fischer-Kowalski, director of the Institute of Social Ecology, he discussed whether he should wrap up his research in the Nicobars and move on. "Many things I was doing, I wouldn't tell my colleagues," Singh says. "It was bad. I was intervening."

CREDITS: SIMRON SINGH

Fond farewell. Simron Singh (*far left*) receives a traditional sendoff—a smearing of coconut oil on his forearm and a garland of banana leaves—from a Nicobarese leader in Pilpillow village on Kamorta Island last April.

Isolate or Engage? Indigenous Islanders Pose Challenge for India

NEW DELHI—Three hours by air from this metropolis are a few societies of ancient lineage: the Nicobarese (see main text) and five other indigenous tribes of the Andaman and Nicobar islands, two of which still pursue a hunter-gatherer lifestyle. The Indian government is grappling with how best to protect these fragile cultures: whether to sharply limit their contact with outsiders or slowly integrate them into modern society.

"The approach we take does not promote complete isolation nor does it advocate complete integration, but a middle ground," says V. R. Rao, director of the Anthropological Survey of India in Kolkata. "Any policy on the Andaman aboriginal groups should allow them a large measure of independence in choosing their own future," adds Sita Venkateswar, a social anthropologist at Massey University in Palmerston, New Zealand, who has studied one tribe, the Onge, for several years.

Remembrance of genes past. Like others of their tribe, this Great Andamanese family has assimilated into modern society.

Indeed, the tribes themselves are largely setting the pace of integration. At one extreme is the Sentinelese of tiny North Sentinel Island in the Andamans. This group, estimated at 100 individuals, may be the last culture in the world maintaining a Stone Age lifestyle, according to the Tribal Welfare Department of the Andaman and Nicobar Administration (ANA). Contacts are virtually nil. "We know about wild tiger numbers in India but don't know the exact population of the Sentinelese," says Vishvajit Pandya, an anthropologist at the Dhirubhai Ambani Institute for Information and Communication Technology in Gandhinagar, who has studied the Onge and the Jarawa.

The Sentinelese prefer isolation. A friendly contact occurred 15 years ago, when an ANA team sailed to the island bearing gifts of cloth, coconuts, and bananas. But that overture was criticized by civil society groups on grounds that the tribe, which had not sought contact, should be left alone to prevent risks such as introduced diseases.

Since 1991, ANA has enforced a hands-off policy toward the Sentinelese. The only exception was a mission to check on how they fared in the 2004 tsunami. When an Indian Air Force helicopter flew over the island, it was greeted with a barrage of arrows and turned back. Then last January, two fishers entered the waters of North Sentinel Island, reportedly to poach crabs. They were allegedly slain and buried in the sand, says Samir Acharya, president of the Society for Andaman and Nicobar Ecology in Port Blair, the Andaman capital. Police exercised restraint by not pressing charges or venturing into Sentinelese territory to retrieve the bodies, Acharya says.

But other tribes are reaching out. The Jarawa, once hostile like the Sentinelese, began to visit ethnic Indian communities in 1998, sometimes seeking medical assistance. Their benign forays pose a challenge for the government: Heightened contact may erode tribal culture, whereas a hands-off approach would be difficult to sustain and justify, particularly when medical aid is sought. The government has since established a health outpost bordering Jarawa settlements.

Prodded by Indian courts, ANA in December 2004 declared the Jarawa reserve "inviolate" and set measures to protect it from further encroachment. Human-rights and environmental groups are not satisfied, however, and petitioned the Supreme Court to force the government to cocoon the Jarawa and other tribes from the outside world; a verdict is expected soon. "On paper, India's policy is … one of the most advanced on isolated peoples anywhere in the world," says Stephen Corry, director of Survival International, a London-based nonprofit organization that promotes the welfare of indigenous peoples and is not involved in the legal action. However, he contends, if authorities do not stiffen enforcement of a no-go zone around Jarawa land, "the Jarawa will not survive."

The other four Andaman and Nicobar tribes—the Onge, Great Andamanese, Shompen, and Nicobarese—are all assimilating into modern society. Last year, a mitochondrial DNA study of Great Andamanese and Onge individuals in *Science* suggested that these peoples could be the oldest surviving human stock in Asia (*Science*, 13 May 2005, p. 996). The tribes are "a gold mine of ancient, undiluted genetic information," says Lalji Singh of the Centre for Cellular and Molecular Biology in Hyderabad, who led the work. But these living links with humanity's past are fraying. The Great Andamanese, who are said to have been 10,000 strong at the end of the 18th century, are down to 20 individuals, and the Onge number only 98. **–PALLAVA BAGLA**

Wolfgang Lutz, an IIASA demographer and leader of IIASA's World Population Program, recalls Singh's quandary. "I remember well discussing with him in 2004 whether this unique culture … should be left entirely alone or whether there is a case for introducing some of the usual development measures such as health care and education," Lutz says. "It was evident from the past experience of many other populations that even the most 'benign' interventions, such as helping to reduce child mortality, in the long run will change the living conditions and therefore also the culture," he says. "My personal preference as a scientist was to restrict our role mostly to observation."

The tsunami, Lutz acknowledges, "made such considerations obsolete."

Cultural annihilation

When the magnitude-9.0 earthquake jolted the Nicobars in the early morning of 26 December 2004, Yusuf was among a group on Nancowry Island who retreated to the beach to keep clear of buildings in case of aftershocks. That's when he says he noticed the sea receding. "I told people to run. There was a hill nearby," Yusuf says. Thanks to Nancowry's favorable geography—it has high ground, and other islands shielded it from the full brunt of the waves—just one person on Nancowry died in the tsunami.

Other islands in the chain were not as fortunate. Some are so low that the waves washed right over them. Trinket, the site of Singh's in-depth study, was carved into three islets. The chief of the tribal council requested Fischer-Kowalski to allow Singh to travel to the Nicobars immediately and stay for a few months. "In this situation, it had little to do with scientific roles but was rather a matter of human reciprocity," Fischer-Kowalski says. Other colleagues preached caution. "We told him, 'The tragedy is

CREDIT: PALLAVA BAGLA

too large; you are only one young expatriate academic with not much influence in India,' " says Martínez-Alier. "We also said, 'Think of yourself.' He did not."

When Singh arrived at the end of January 2005, he was confronted with utter devastation. Most survivors were living under tarpaulin shelters. And their cultural heritage had been obliterated. All but two *kareau*, carved wooden effigies bearing ancestral bones, were gone. "We can bring back pigs," Yusuf says. "But the bones are lost."

Exposed to new ideas—and the elements. Nicobarese children in a makeshift school at a relief camp on Kamorta Island.

To the rescue

Back in Vienna, Fischer-Kowalski was rallying institutional support. Klagenfurt University, which oversees her institute, spearheaded a fundraising drive for reconstruction. Under the auspices of Caritas Austria, a Catholic relief agency, and Universal Music Group, which donated the proceeds from sales of a CD recorded to support tsunami victims, the university set up the Sustainable Indigenous Futures (SIF) Fund to support self-rehabilitation efforts of indigenous peoples affected by the tsunami. And the Austrian Science Fund stepped up to bankroll a scientific assessment of sustainable development in the Nicobars.

Many colleagues applaud Singh for taking on these responsibilities. "As I see it now, there is no dilemma left for the scientist at this point," says Lutz. Singh, he says, is "bringing the best scientific information available to the attention of the local decision-makers" as they plan for reconstruction of homes and other infrastructure. "Disasters happen daily to some of the peoples we study," adds Martínez-Alier. "I ask myself why it does not happen more often that social scientists turn into advocates."

Last September, Singh's institute and the SIF Fund hosted six Nicobarese tribal leaders. Colleagues took them to Austrian villages rebuilt after the disastrous floods of 2003. "These were villages with 900 years of history," says Singh. "We wanted to show them that no matter how difficult the ordeal, you don't ever have to give up your culture."

To help reconstitute the islands' cultural heritage, Singh is drawing on an extraordinary resource. The Museum of Ethnology Vienna has more than 200 Nicobarese artifacts collected in the 19th century—in the wake of the tsunami, one of the largest collections in the world. Some objects will be lent to the Nicobarese and copies of others manufactured. "These remind us of our way of life and will help us preserve it," Yusuf says.

Other momentous changes are on the horizon. The Indian government, pending approval from the defense establishment, may open the Nicobars to expanded trade and tourism. The Nicobarese "aren't saying they want to live like museum pieces," Singh says. But they don't want to lose their identity by integrating fully with the outside world. "What Simron is doing that's so special," says Brian Durrans, deputy keeper in the British Museum's Asia Department, "is a combination of refusing a patronizing 'isolationist' option while encouraging the Nicobarese to become their own advocates. It's a pretty inspiring approach in conditions of sudden catastrophe."

Singh is also advising the Nicobarese on how to diversify their economy, which had been based almost entirely on swapping or selling coconuts and coconut products for goods from ethnic Indians. Before the tsunami, Singh says, the Nicobarese were "not prepared" to shift away from a reliance on coconuts. Now they understand that it will take years to restore the palm groves. In the meantime, they expect to derive income from fishing and garden plots.

All the more valuable. One of only two *kareau*, effigies bearing ancestral bones, that was not swept away by the tsunami.

CREDITS: SIMRON SINGH

Singh, who returned to Vienna in May after a 2-month stint in the Nicobars, says the latest challenge is helping the islanders learn the value of money. Recently, each family received cash compensation from the Indian government. The money has been burning a hole through the tribal pocket, as islanders pay two or three times the going rate for everything from motorcycles to DVD players. "The accumulation of capital is rare, if not unknown," Singh says. "It's the major problem at the moment."

The jarring transitions have driven many young Nicobarese in search of a better life to Port Blair, the relatively developed capital of the Andaman Islands. "I fear that the younger generation might turn their backs on us," Yusuf says. Samir Acharya, president of the Society for Andaman and Nicobar Ecology in Port Blair, believes a mass exodus is unlikely. He says that conversations with a few dozen young Nicobarese in Port Blair suggest they are "all likely to go back to their respective villages." Nevertheless, most experts agree that the Nicobarese culture is hanging by a thread. The situation, says Gooch, "is really grim."

Now Singh has reached another crossroads. At the moment, he does not know whether to return to the role of detached observer, turn toward advising the Indian authorities, or continue with reconstruction projects. Singh's colleagues are confident he will choose a noble path. "He'll do the right thing," says Gooch—as a scientist and as an advocate.

–RICHARD STONE

Published by AAAS

Following their noses. USDA's Don Barnard works with a chamber that allows mosquitoes to follow or avoid certain smells.

MEDICAL ENTOMOLOGY

Keeping the Bugs at Bay

Public health money gives a boost to the untidy science of crafting a better insect repellent

Andrew Fowles, a high-school security guard in Davis, California, is a part-time grunt in the war on biting insects. For $30 an hour, he has served as a test subject from Sacramento to the Florida Keys, dousing himself with experimental bug repellents, then thrusting an arm into a mosquito-filled cage or marching bare-legged through bug-swarmed wetlands. He once got about 1500 mosquito bites on his calf—two bites per second. "I'm not squeamish. I like to contribute to society," he says.

Bug-repellent research, long waged on primitive levels and with mixed results, may be looking up. The developed world once viewed repellents chiefly as products for backyard comfort, but with the spread of insect-borne agents such as West Nile virus, they have become a public-health issue. There is also growing recognition that vector-borne diseases routinely devastate the developing world; in Africa alone, 800,000 children die from malaria every year. Expensive research into vaccines, medicines, and genetic modifications to wild insects has so far delivered little, so many scientists are going back to basics: insecticides, bed netting, and repellents.

With new funding initiatives, biologists are using more sophisticated methods to improve understanding of insects' finely evolved olfactory systems and to find substances that might disrupt them. Researchers have uncovered possible new repellents in nature and in the lab, and a few have already made it to market; others may be on the way. "If we can really understand how insects find us, we can figure out how to prevent them from finding us," says Yale University molecular neurobiologist John Carlson, part of a new three-continent team.

Free lunch. Mosquitoes gorge themselves on an artificial food source in the lab.

Trying to beat DEET

Cultures from ancient Egyptians to modern U.S. Southerners have used remedies as diverse as snakeskin and cow parsnip to keep away pests. But so far the Holy Grail—a compound that bugs hate and humans find pleasant—has been elusive. The central weapon in the modern arsenal is N,N-diethyl-meta-toluamide, or DEET; invented in 1953, it smells evil, melts plastic, and is perceived by many people to be poisonous. No one even quite knows how it works. At 30% concentrations, DEET drives away a broad spectrum of bugs for up to 8 hours, including *Culex pipiens* mosquitoes, which carry West Nile virus. However, many commercial products carry as little as 7% DEET because it's so unpleasant, and even high concentrations fall short against *Anopheles* mosquitoes, which spread malaria, and against ticks, which spread Lyme disease and Rocky Mountain spotted fever. The U.S. military has found that in the fierce heat of Iraq, DEET evaporates quickly. In the current war, biting sandflies have inflicted close to 2000 cases of leishmaniasis, which causes debilitating skin lesions.

Psychology also plays a role. The U.S. Department of Defense recently revived work on new repellents after finding that soldiers often toss out green tubes of military-issue DEET because they fear it is toxic and hate its sticky feel. Some unknowingly replace it with the same stuff in a more colorful package aimed at civilians, says Lt. Col. Mustapha Debboun, a medical entomologist at the Army Medical Center and School in Fort Sam Houston, Texas.

DEET actually has a good safety record, asserts Debboun, who heads the armed forces' repellent committee and is co-editor of the forthcoming book *Repellents*, the first overall look at the field. The compound is used hundreds of millions of times worldwide each year, with only about 50 known reports of severe dermatitis or seizures since it came into use—problems usually attributed to gross overuse, if they can be clearly connected to the product at all. "But if soldiers are afraid of it, we need to look for alternatives," says Debboun.

Some researchers are turning to folk remedies—and finding a few that may work, at least to a degree. The U.S. Department of Agriculture (USDA), which has traditionally led American research at the relatively modest funding rate of about $1 million per year, has shown that many plant oils, including clove, peppermint, geranium, and catnip, may repel mosquitoes. Botanists have long known that plants produce compounds to combat pests that might eat them; entomologists speculate that there may be a chemical spillover effect on those that eat us.

For example, last year Stoneville, Mississippi–based USDA botanist Charles Bryson showed that his grandfather was right when he shoved American beautyberry leaves under his draft animals' harnesses to repel biting flies. Bryson and colleagues reported in the *Journal of Agricultural and Food Chemistry* that they have isolated five com-

CREDITS (TOP TO BOTTOM): ARS/USDA; AIJUN ZHANG; (PEST IMAGES) DAVID LIEBMAN; ARS/USDA; SCOTT BAUER/USDA

Published by AAAS

pounds that repel bugs including *Aedes aegypti*, the mosquito species that carries yellow fever. They have patented one substance, called callicarpenal, and are looking for a commercial partner for development.

Animals may also employ natural repellents. Recently, Paul Weldon, an animal behaviorist at the Smithsonian Institution in Washington, D.C., observed that some monkey species rub themselves with millipedes that secrete benzoquinones, compounds shown in the lab to repel insects. He reported in the journal *Naturwissenschaften* in January that coatis sense benzoquinones and roll millipedes around in an apparent effort to draw them out. Weldon says that many other mammals and birds anoint themselves with plants or other items, many of which show some repellent effect.

Other creatures, from giraffes to gaurs, may produce their own repellents. Auklets on arctic islands exude aldehydes that drive off ticks and mosquitoes and kill lice in lab experiments, according to a paper last year in the *Journal of Medical Entomology* by ecologist Hector Douglas of the University of Alaska, Fairbanks. Repellents are not trivial for the birds; the insects are so vicious that they can threaten breeding colonies' existence. Douglas says that during buggy times, one can smell the citrusy aldehydes a kilometer away from colonies.

But there are many obstacles between identifying natural repellents and deriving marketable products from them. A widely cited 2003 study in *The New England Journal of Medicine* by Chapel Hill, North Carolina, dermatologist Mark Fradin shows that many "botanicals," including popular citronella formulations such as Avon Skin-So-Soft, work—but only for 3 to 20 minutes. Part of the problem is that the compounds are usually quite volatile and evaporate unless constantly replaced. Some natural oils will repel bugs for up to 2.5 hours, but only if they are distilled down to something near the pure stuff, which is usually stinky and toxic, says Don Barnard, a USDA entomologist in Gainesville, Florida. Aldehydes, for example, smell nice in low doses but are hard to bear once purified. High doses also may corrode people's mucous membranes and livers. "I know one lady who put 100% clove oil on her face, and she got horribly burned; she must have been pretty desperate to go organic," says Barnard.

All the same, researchers continue to hunt for natural repellents, including those produced by humans. Ulrich Bernier, a chemist at USDA's Agricultural Research Service in Gainesville, has identified about 275 substances in human sweat, including aldehydes similar to those produced by auklets, along with ketones, fatty acids, and ammonia. Human breath contains carbon dioxide—long known as the crudest insect attractant—but also may contain substances similar to those in sweat. Bernier says many, such as lactic acid, act as attractants at low concentrations but become repellents when distilled, combined in certain proportions with other emanations, or modified into slightly larger molecules. Everyone sweats out the same chemicals, but the concentrations and proportions may vary widely among individuals—possibly the key to why bugs eat some folks alive but leave others alone. Bernier has already patented several chemical leads derived from his research.

One group at the United Kingdom's government-funded Rothamsted Research has combined human and animal research. Biochemist John Pickett has shown that some cattle exude powerful compounds that keep biting flies off themselves and nearby animals. The tendency to produce the substances appears to be inherited through the mother, he says. Rothamsted chemical ecologist James Logan says humans produce similar compounds; in as-yet-unpublished work, he has isolated what may be human emanations that work against *Aedes aegypti* mosquitoes and Scottish no-see-ums. With a new $500,000 grant from a government fund for new business enterprises, the group hopes to do field trials in South America and Africa over the next 2 years.

Strongest weapon. The military issues DEET, the most widely used repellent, but not all soldiers like it.

On guard. Auklets produce natural repellents.

The discovery process is slowed, however, not only by the complexity of how aromas interact with one another, but also by the varying tastes of different insect species, strains within species, and even individual insects. Says Bernier: "We're just beginning to understand: Insects are like picky wine drinkers. They like something or not, based on very subtle combinations of qualities." What repels a mosquito may have no effect on a chigger, so a versatile repellent may demand a chemical cocktail.

Testing troubles

Repellents' exact modes of operation remain mysterious, and this has also hindered research; most substances now in use were identified through fortuitous observations or laborious random trials. A few months ago, Jerome Klun, a USDA entomologist based in Beltsville, Maryland, finally showed in the *Journal of Medical Entomology* that vapors from DEET and some other substances function mainly by traveling to insects' olfactory receptors, as opposed to affecting the insects when they touch skin—a mechanism long suspected but never proven till now. However, Klun says it is still unclear whether the substances create an unpleasant sensation for insects, mask attractive odors, or work in some other way.

To discover repellents without knowing why they work, one new Army initiative uses computers to identify molecules similar to existing repellents, then synthesizes and tests the compounds. In unpublished research, Army biochemist Raj K. Gupta, based at the Walter Reed Army Institute of Research in Silver Spring, Maryland, has made a sort of repellent template by identifying various electrical and structural qualities common to existing products. From this, a team has come up with at least four compounds now under investigation.

CREDITS (TOP TO BOTTOM): HECTOR DOUGLAS; LT. COL. MUSTAPHA DEBBOUN/ACADEMY OF HEALTH SCIENCES

Sacrificial arm. Testers bravely offer their flesh to caged mosquitoes.

But even if compounds prove interesting, documenting their effectiveness is often complicated. For example, a few new repellents are now on the market, but it's hard to compare them to each other, or to DEET. One substance is para-menthane-3,8-diol (PMD), commonly known as oil of lemon eucalyptus. Derived from the Australian lemon-scented gum tree and marketed in China for years, it appeared last year in the United States under the brand name Repel. In some tests, 20% PMD appears to be nearly as effective as most of the 240-some standard DEET products—perhaps better, says Scott Carroll, the University of California, Davis, entomologist who sent out Fowles to test the stuff. (Both Fowles and Carroll now swear by it.) Other studies, however, suggest that PMD is weaker than DEET. One problem is that there are 3000 species of mosquitoes alone, and most repellents work better on some than on others. But many tests using caged mosquitoes are conducted using only one or two species. And in the field, insects' tastes may vary dramatically by concentration of the repellent, time of day, temperature, humidity, individual volunteers, and insect species themselves—variables that few researchers try to control for.

Two other promising repellents are newly available in the United States: Picaridin, a synthetic derivative of pepper marketed under the name Cutter Advanced, and IR3535, a derivative of a single amino acid, β-alanine, now sold in a beefed-up version of Skin-So-Soft. However, research on them suffers similar complications. "What's best? It sounds like a simple question, but there really is no easy answer," says Robert Novak, a medical entomologist at the Illinois Natural History Survey.

Testers face another hurdle: With the spread of West Nile and other diseases to new regions, field tests against wild insects may expose human volunteers to infection. Researchers are getting creative. Instead of letting volunteers get bitten, USDA scientists now trail them through the buggy Florida Everglades with portable aspirators to suck off any mosquitoes that land and probe; most insects carry through with feeding once they start this sequence, so these events are counted as bites. For lab work, Klun and Debboun have invented the so-called K&D module: a plastic box with multiple mosquito-filled chambers pressed against a volunteer's thigh, which allows researchers to test out several repellents at once using lab-cultivated, disease-free mosquitoes.

One improved system does away with the humans altogether by substituting a skinlike membrane filled with human blood, heated to human body temperature and treated with repellents. Klun says human volunteers will always be needed for final-stage field tests, but the new technology will minimize risk during the lengthy screening that usually precedes such tests.

Perhaps the most sophisticated repellent effort, aimed at stages from basic biology to field tests to marketing, is a 5-year project begun last September and funded by the Bill and Melinda Gates Foundation. The $8.5 million budget is comparable to that of USDA and the U.S. military combined. Following their 2001 identification of the genes for the 80-some olfactory receptors used by malaria-carrying *Anopheles gambiae* mosquitoes, two labs in the United States have begun running tests with hundreds of substances to observe which cause receptors either to fire or become blocked. First, researchers transplant genes for the receptors into other biological systems that are faster and easier to work with than mosquitoes, such as frogs' eggs or fruit flies. Then they test whether various substances affect electrical currents flowing from the receptors, presumably an indicator of firing.

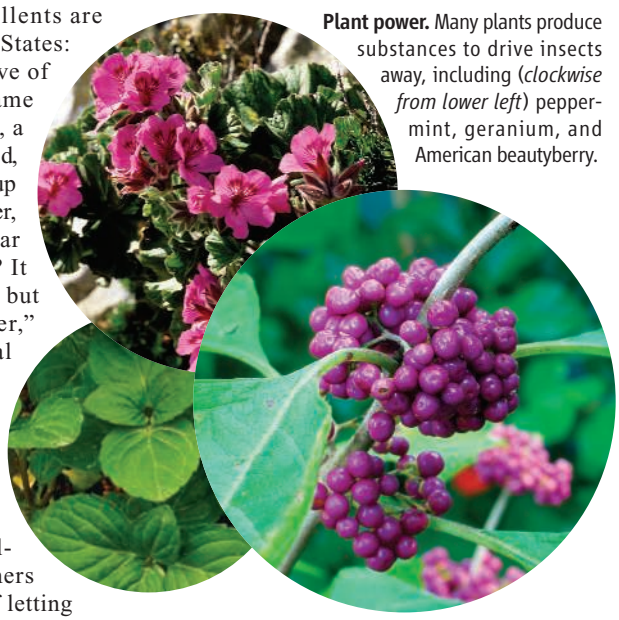
Yale's Carlson has been working in part with compounds taken from human sweat and tested in fruit flies that express mosquito receptors. He has already sent a half-dozen candidate compounds on to the next stage: the lab of behavioral entomologist Willem Takken at the University of Wageningen in the Netherlands. Takken's job is to test the effects on mosquitoes' antennae and behavior. After again measuring whether receptors fire, this time in the mosquitoes' antennae, he employs wind-tunnel–like olfactometers, in which insects can choose to follow or avoid certain paths from which different smells emanate.

Compounds that show promise will be sent to the Ifakara Health Research and Development Centre in Tanzania, where a greenhouselike biosphere about half the size of a football field is now under construction. Here, human volunteers will test candidate repellents against free-flying mosquitoes in a semicontrolled environment within the next year or so. Top picks will go to Gambia, where researchers will do full field tests in villages—then, hopefully, to market by 2011.

Takken, Carlson, and Laurence J. Zwiebel of Vanderbilt University in Nashville, Tennessee, emphasize that they are looking not just for repellents but also attractants, following other research on what draws insects to their targets (*Science*, 4 October 2002, p. 176). They hope to develop a "push-pull" system that will employ repellents to drive mosquitoes away from people's beds or other feeding spots and attractants to draw them into insecticide-laced traps in distant parts of houses or villages. Their repellent would not be applied to skin but rather emitted from a simple device that would suffuse a room or outdoor area—a highly evolved version of the mosquito coil, which has long used various substances, although with only fair to middling effect. Takken says that because attractants and repellents seem related, and no single compound works on all bugs, the solutions are likely to be complex. He has already identified one apparently powerful new repellent that he hopes to publish next month. Zwiebel says the key will be to keep the system cheap and to use sales of repellents in richer countries to subsidize use in poorer ones. "The Third World can't afford what we spend to protect our backyard barbecue," says Zwiebel. "But they really need repellents."

–KEVIN KRAJICK

Kevin Krajick is a New York City–based writer who is unattractive to bugs.



Plant power. Many plants produce substances to drive insects away, including (*clockwise from lower left*) peppermint, geranium, and American beautyberry.

CREDITS (TOP): ARS/USDA; (FROM BOTTOM LEFT, CLOCKWISE) MICHAEL THOMPSON/USDA; THOMAS ELIAS/USDA; J. S. PETERSON/USDA

ASTROPHYSICS

A Towering Physicist's Legacy Faces a Threatening Future

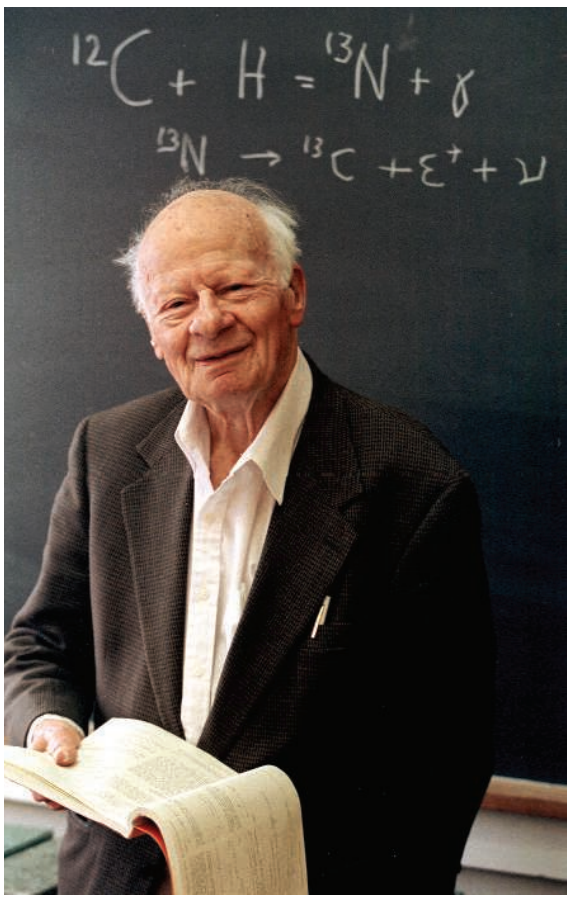
On the centennial of Hans Bethe's birth, his successors worry that cuts in long-planned projects will discourage the next generation of brilliant minds

ITHACA, NEW YORK—The language of astrophysics sizzles with alpha particles and gamma rays. There's a heavy dose of beta as well—Hans Bethe, that is, a giant of 20th century science, whose prowess in nuclear physics led to his fascination with combustion in deep space.

Bethe probed astrophysics at its purest levels right up until his death last year at age 98 (*Science*, 8 April 2005, p. 219). At a recent meeting* here, speakers fondly recalled his influence in the region where nuclear physics and astrophysics fuse, from neutrinos to supernovae, and ordinary stars to neutron stars. They also laid out key mysteries that still tantalize scientists: How do giant stars explode and forge the elements around us? What happens when neutron stars or black holes crash? And what is the nature of the dark matter and dark energy that suffuse space?

But as Bethe's scientific descendents marked what would have been his 100th birthday on 2 July, they worry about their ability to address such questions anytime soon. Cuts in the science program at NASA have cast a pall over missions designed to turn the cosmos into a high-precision physics laboratory. The damage to Bethe's legacy could be serious, they warn.

"In the worst-case scenario, the young people we need may feel hopelessness," says Saul Teukolsky, chair of the physics department at Cornell University—Bethe's academic home for 70 years. "They may not enter the field at all."



A rare spark. Hans Bethe, shown at age 90, calculated how stars burn—including the "carbon cycle" in background.

CREDIT: MICHAEL OKONIEWSKI

Bomb physics, near and far

Astrophysics was the alpha and omega of Bethe's long career. Although his fame stems from leading the theoretical division at Los Alamos, New Mexico, during the Manhattan Project, and his tireless advocacy of arms control once World War II was over, astronomers and physicists revere him as "the guy who figured out how the sun works," says astrophysicist Michael Turner of the University of Chicago, Illinois. "You don't need a better legacy than that."

Born in Strasbourg, Germany, and educated at universities in Frankfurt and Munich, the young Bethe spelled out the details of the proton-proton reaction that propels hydrogen fusion in the cores of modest stars like our sun. In the late 1930s, he was the first to describe a separate fusion cycle involving carbon, nitrogen, and oxygen atoms, which powers massive stars during their short lives. In 1967, Bethe received the Nobel Prize in physics for that work.

In the decades after World War II, Bethe's research focused on the theory of nuclear matter and atomic physics. A highlight was a calculation of a subtle shift in the energy levels of electrons in excited hydrogen atoms. That three-page paper, written on the train between New York City and Ithaca, set the stage for modern quantum electrodynamics.

Later in life, however, two catalysts drew Bethe tirelessly back into astrophysics. The 1967 discovery of pulsars—flashing deep-space beacons that Bethe's Cornell colleague Thomas Gold explained as spinning neutron stars—sparked Bethe's intense desire to understand the properties of superdense states of matter. Drawing from his deep well of nuclear physics, Bethe and colleagues wrote papers on the internal structures of neutron stars. They derived a likely radius of 10 kilometers, a figure still in vogue.

Soon after Bethe "retired" in 1976, his friend Gerald Brown of the State University of New York, Stony Brook, piqued his interest with a challenge to work out the nature of supernovae. The two scientists spent much of the next 3 decades pondering how giant stars blow up, their prodigious outbursts of neutrinos, and binary systems of neutron stars and black holes.

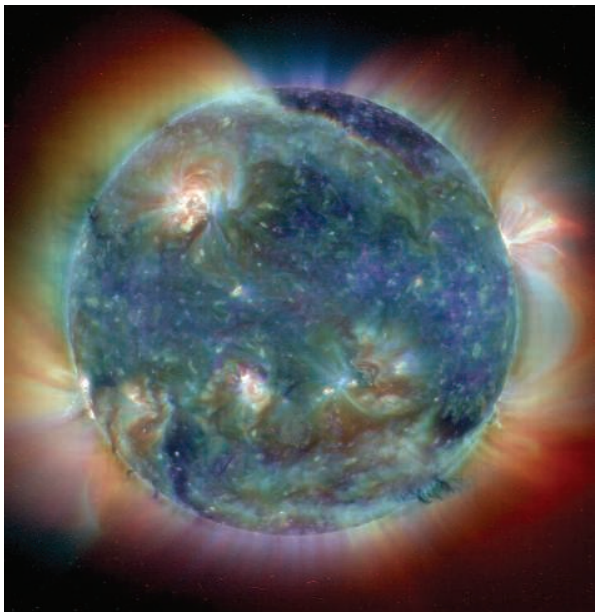
These topics followed logically from Bethe's work on the atom bomb, says astrophysicist Stan Woosley of the University of California, Santa Cruz. "Stars are gravitationally confined thermonuclear reactors," and their demise is bomb physics on the grandest scale, Woosley says. "And Hans was really interested in the birth of the elements, especially uranium," he adds with a smile.

Bethe wanted to find the essence of why a dying star's core implodes. His key contribution, says Woosley, was to consider the star's entropy. As a star runs out of fuel and fuses heavier elements up to iron, Bethe found, the outer layers grow disordered while entropy declines at the blazing core. "Hans liked to say [the core] had the entropy of an ice cube, even though it was 10 million times hotter than hell," Woosley recalls. Bethe calculated that when the core collapses, it has too little entropy for iron nuclei to break up. Instead, they compress into the extraordinary densities of neutron stars.

That collapse ignites an outward shock wave, which the great mass of the star quickly snuffs out. Bethe believed neutrinos emitted by the newborn neutron star would relaunch the shock wave and drive the supernova blast, a scenario he published in his 80s with astrophysicist James Wilson of Lawrence Livermore National Laboratory in California. The verdict is still out; the best computer models have yet to blow up a simulated star in a convincing way.

Solar neutrinos also captivated Bethe. With the late John Bahcall of the Institute for Advanced Study in Princeton, New Jersey, Bethe helped explain why underground detectors on Earth observed only a fraction of the

* Bethe Centennial Symposium on Astrophysics, Cornell University, 2–3 June.



Ablaze with energy. Neutrinos from our sun drew Bethe's focus late in his career.

neutrinos predicted to stream from the sun's core. Confirmation came from Canada's Sudbury Neutrino Observatory in 2001: The particles have minuscule masses and oscillate among different "flavors." That behavior, as Bethe and Bahcall foresaw a decade earlier, arises from unknown physics beyond today's standard theory.

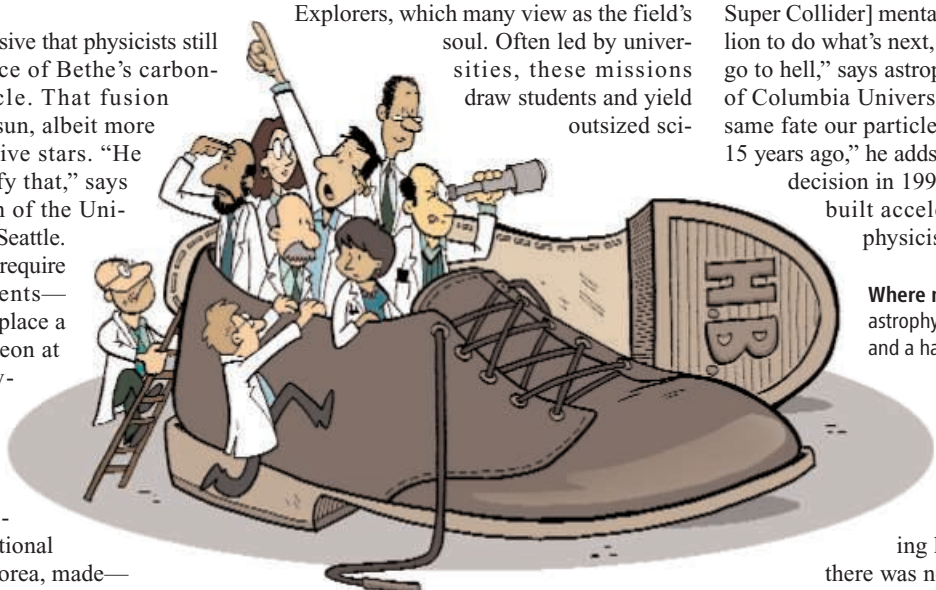
Neutrinos are so elusive that physicists still have no direct evidence of Bethe's carbon-nitrogen-oxygen cycle. That fusion should happen in our sun, albeit more sedately than in massive stars. "He would want us to verify that," says physicist Wick Haxton of the University of Washington, Seattle. Doing so, however, will require sensitive new experiments—such as a proposal to place a 130-ton vat of liquid neon at Sudbury to spot low-energy neutrinos.

Bethe also did not live to see a test of a claim that he, Gerald Brown, and Chang-Hwan Lee of Pusan National University in Busan, Korea, made—after Bethe turned 90—that binary systems containing two black holes should be 20 times as abundant as systems with one black hole and one neutron star. Their prevalence would be good news for the Laser Interferometer Gravitational-Wave Observatory (LIGO), which seeks the space-rippling disturbances caused by the mergers of such binaries. Two black holes should make a more violent "splash" in the gravitational pond of space, says theorist Kip Thorne of the California Institute of Technology (Caltech) in Pasadena, who motivated Bethe to probe the issue.

An endangered generation

The Laser Interferometer Space Antenna, long under development with the European Space Agency as a sensitive partner to LIGO, is one of three major space science missions planned by the astrophysics community in the next decade to peer further into Bethe's realm. The other projects are the four telescopes of Constellation-X, a high-resolution successor to the Chandra X-ray Observatory; and NASA's share of the Joint Dark Energy Mission, an effort with the U.S. Department of Energy to chart the weird speeding-up of the universe's growth and determine its cause. But funding prospects are dim (*Science*, 17 March, p. 1540). A tight NASA budget, combined with massive cost overruns and a huge backlog of proposed projects, has left them competing for what could be only one new NASA start for a major astrophysics mission in the next 3 years.

The pain spreads to NASA's low-cost Explorers, which many view as the field's soul. Often led by universities, these missions draw students and yield outsized scientific results, says astrophysicist Roger Blandford, director of the Kavli Institute for Particle Astrophysics and Cosmology at Stanford University in Palo Alto, California. "You optimize the use of finite resources most efficiently by putting them into Explorer programs," Blandford says. "Instead, NASA has been starving them to death," including cancellation in February of the on-budget NuSTAR mission to image high-energy x-rays.



Where next? In the post-Bethe era, astrophysicists face tough choices—and a hard act to follow.

Some researchers at the meeting pinned these sacrifices on NASA's decision, after much political intervention, to fund a repair mission for the Hubble Space Telescope and to push ahead with its planned successor, the costly James Webb Space Telescope (JWST). "We could just see [Hubble] would be a $2 billion drag on the program," says Turner, who recently completed a 3-year stint as head of mathematical and physical sciences at the National Science Foundation. Given all of NASA's other priorities, Turner says, moving forward with the repair without reconsidering its value within the entire suite of missions "was stupidity on stilts."

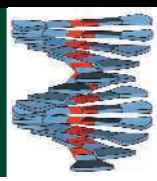
But other scientists believe that prolonging Hubble's life didn't automatically take money away from other programs because space science isn't a zero-sum game. Even with no servicing mission, "it's not clear that money would be made available for Explorers, or Constellation-X, or anything," says astronomer Robert Kirshner of the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts. He also notes that Bahcall led a National Academies' committee that in 2003 came out strongly in favor of at least one more NASA repair flight to the telescope, after extensive community input.

More vexing for the long-term health of astrophysics is the ballooning cost of JWST, now pegged at $4.5 billion. "I worry that we've gotten ourselves into the SSC [Superconducting Super Collider] mentality, that we need $5 billion to do what's next, and everything else can go to hell," says astrophysicist David Helfand of Columbia University. "We may suffer the same fate our particle-physics colleagues did 15 years ago," he adds, referring to Congress's decision in 1993 to cancel the partially built accelerator in Texas. Astrophysicist Shri Kulkarni of Caltech paints the situation bluntly: "Is a single mission worth the rest of astronomy?"

No one at the meeting had a good answer, and there was no consensus on how the community might gain the necessary political support for its priorities. Indeed, the room seemed infused with a wistfulness that Bethe couldn't be there to rally his colleagues in their time of need. "The scope of problems he could solve pretty much had no limit," Brown wrote last year in *Physics Today*, recalling his struggle to keep up with a friend 20 years his senior. "In that sense, I think [Bethe] was the most powerful scientist of the 20th century."

–ROBERT IRION

CREDITS (TOP TO BOTTOM): NASA/ESA; ILLUSTRATION: PAT N. LEWIS



LETTERS

edited by Etta Kavanagh

Public Access Failure at PubMed

THE NIH PUBLIC ACCESS POLICY REQUESTS THAT NIH-SUPPORTED INVESTIGATORS SUBMIT FINAL peer-reviewed primary research manuscripts to the PubMed Central database (PMC) upon acceptance for publication (*1*). The policy went into effect 2 May 2005. As of January 2006, only approximately 3.8% of NIH-funded research papers published after 1 May 2005 had been submitted to the PMC repository (*2*).

Low compliance only tells part of the story. More than half of the manuscripts available on PMC were published before 2 May 2005 (*3*). Many reviews and commentaries, which fall outside of the scope of the request, and papers inappropriately made publicly available before the publisher's public access embargo were also found in the database. This suggests either wide misunderstanding of the policy or deliberate submission of papers falling outside the scope of the database.

The policy also allows posting of papers that differ significantly from the final published version, which has the potential to create intellectual property issues as each public disclosure of the research represents prior art in the eyes of the law. Also, there is no dedicated system to guarantee that corrections made after publication, which can be significant, are made to the author-submitted paper.

By NIH estimates, if only half of the eligible papers are submitted to the database, the cost would reach $2 million per year, or $62 per paper (*2*). Without a mandatory policy, however, submission of half of all eligible papers is unlikely. The NIH already provides close to $30 million annually to cover publication costs. As the policy expands, archiving could cost an additional $3 million (*4*).

The submission rate over the course of 2005 varied little. Submissions have increased significantly since then, but are still not approaching full compliance (*3*, *5*). Both internal and external warnings that, if voluntary, the program would fail were outweighed by the NIH's desire to allay the concerns of some publishers and those advocating public access policies.

There is some good news, though. Authors publishing in some of the more influential journals in biomedical research seem to have a higher compliance rate than the estimated average (*3*). There is no obvious link between journal cooperation and author participation or any clear explanation for the journal-to-journal variability, but it is still a positive sign for PMC.

> **"NIH's faltering experience** so far indicates that public access policies must be mandatory and curated if they are to have any chance of success."
>
> —Stebbins ***et al.***

Notably, we still lack a demonstrated desire by the general public for access to primary research papers, leaving the true public value of the repository an open question on a backdrop of a disinterested scientific community and angry publishers and societies. The public access movement is spreading quickly, nonetheless.

Senators Cornyn (R-TX) and Lieberman (D-CT) recently introduced the Federal Research Public Access Act (S.2695), which imposes a mandatory public access policy on publications resulting from research funded by all federal agencies with extramural research expenditures over $100 million. Not surprisingly, the bill has drawn criticism from many publishers and societies, some of whom feel that it unfairly places scientists between funding agencies and publishers. An April European Commission report recommends that funding agencies promote public access to research publications and suggests that agencies make compulsory deposition a condition for funding (*6*). Research Councils UK released a draft open access policy last June that called for a mandatory policy at the earliest opportunity (*7*).

NIH's faltering experience so far indicates that public access policies must be mandatory and curated if they are to have any chance of success. It would also be wise for there to be a real demonstration of public desire or need before we expand it to other agencies. Unfortunately, this experiment has cost taxpayers money and the NIH credibility.

MICHAEL STEBBINS,[1]* ERICA DAVIS,[2] LUCAS ROYLAND,[1] GARTRELL WHITE[1]

[1]Federation of American Scientists, 1717 K Street, NW, Suite 209, Washington, DC 20036, USA. [2]Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA.

*To whom correspondence should be addressed. E-mail: mstebbins@fas.org

References

1. *Fed. Regist.* **70** (26), 6891 (2005).
2. See http://publicaccess.nih.gov/Final_Report_20060201.pdf.
3. See http://www.fas.org/pmc.
4. E. Zerhouni, *Science* **306**, 1895 (2004).
5. Statistics can be found at http://www.nihms.nih.gov.
6. See http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf.
7. See http://www.rcuk.ac.uk/access/index.asp.

Connectivity in Marine Protected Areas

MARINE PROTECTED AREAS (MPAS) ARE A PROMising tool for many problems, from biodiversity conservation to fisheries management (*1*). Their success depends on connectivity among protected areas and spillover into unprotected areas. In their Report "Scaling of connectivity in marine populations" (27 Jan., p. 522), R. K. Cowen *et al.* integrated key ecological factors important in the design of MPAs to show lower connectivity—i.e., reduced larval dispersal between and greater larval retention within reef systems—than previously predicted among Caribbean reefs. In the accompanying Perspective "Staying connected in a turbulent world" (27 Jan., p. 480), R. S. Steneck noted that connectivity will be further reduced by habitat fragmentation and overfishing. The solution Steneck noted, that marine resource managers must protect their reefs on a local scale, makes considerable ecological sense; practice and theory have shown that this will increase the abundance and size of fish, thereby promoting connectivity and spillover (*2*, *3*). A negative, evolutionary impact of local protection on connectivity may offset these advantages, however.

Published by AAAS

As commercial or recreational fishing pressure intensifies outside reserves (*4*), local protection can select for decreased dispersal distance (and increased local recruitment) (*1*), thus accelerating the fragmentation of reefs via rapid evolution of life-history characteristics and potentially reducing the genetic capacity of resident organisms to respond to future environmental change. Analogous effects may limit spillover, and because dispersal distance is likely genetically correlated to larval development and size at maturation (*5*), the effects of MPAs on life-history evolution could be synergistic. As with the known effects of selective harvesting on the evolution of fish life histories (*6*), notably age and size at maturation (*2*, *7*), these changes could be rapid enough to be measured with existing methods (*8*, *9*). The results could be used to construct networks of MPAs, perhaps of various sizes and spacing (*10*, *11*), designed to maintain ecosystem function on evolutionary time scales. These networks may implicitly also be well suited to protecting different life-history stages and the diverse life histories of the varied organisms that they harbor.

MICHAEL N. DAWSON,[1]* RICHARD K. GROSBERG,[1] LOUIS W. BOTSFORD[2]

[1]Section of Evolution and Ecology, College of Biological Sciences, [2]Department of Wildlife, Fish and Conservation Biology, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA.

*To whom correspondence should be addressed. E-mail: mndawson@ucdavis.edu

References

1. L. W. Botsford, A. Hastings, S. D. Gaines, *Ecol. Lett.* **4**, 144 (2001).
2. C. M. Roberts, J. A. Bohnsack, F. Gell, J. P. Hawkins, R. Goodridge, *Science* **294**, 1920 (2001).
3. M. L. Baskett, S. A. Levin, S. D. Gaines, J. Dushoff, *Ecol. Applic.* **15**, 882 (2005).
4. F. C. Coleman, W. F. Figueira, J. S. Ueland, L. B. Crowder, *Science* **305**, 1958 (2004).
5. L. A. Levin, J. Zhu, E. Creed, *Evolution* **45**, 380 (1991).
6. C. A. Stockwell, A. P. Hendry, M. T. Kinnison, *Trends Ecol. Evol.* **18**, 94 (2003).
7. D. O. Conover, S. B. Munch, *Science* **297**, 94 (2002).
8. S. E. Swearer, J. E. Caselle, D. W. Lea, R. R. Warner, *Nature* **402**, 799 (1999).
9. G. P. Jones, S. Planes, S. R. Thorrold, *Curr. Biol.* **15**, 1314 (2005).
10. S. R. Palumbi, *Annu. Rev. Environ. Resour.* **29**, 31 (2004).
11. D. M. Kaplan, L. W. Botsford, *Can. J. Fish. Aquat. Sci.* **62**, 905 (2005).

Letters to the Editor

Letters (~300 words) discuss material published in *Science* in the previous 6 months or issues of general interest. They can be submitted through the Web (www.submit2science.org) or by regular mail (1200 New York Ave., NW, Washington, DC 20005, USA). Letters are not acknowledged upon receipt, nor are authors generally consulted before publication. Whether published in full or in part, letters are subject to editing for clarity and space.

Response

DAWSON *ET AL.* RAISE AN INTERESTING AND important point about possible negative evolutionary consequences of managing metapopulations of coral reef–dwelling fish as they become increasingly isolated due to overfishing and habitat loss. This could potentially be problematic should managers elect to establish small reserves on the basis of relatively small, ecologically relevant dispersal distances. To this point, Cowen *et al.*'s original premise was that the tails of the dispersal kernel, which encompass maximum dispersal distances, were genetically relevant (i.e., genetic exchange would occur over large distances). Further, their finding that some subsidy is important for the maintenance of almost all populations and that there are regional patterns in connectivity that map onto genetic data suggests the importance of larger-scale genetic population connectivity, in addition to small-scale ecological connectivity. Although recent evidence (*1*) may counter the concern raised by Dawson *et al.*, as genetic diversity actually increased within reserves, the authors nonetheless also warn that "local measures are insufficient [for genetic

exchange] when the scale of connections encompasses large areas of territory." In fact, these points suggest the importance of a network of reserves (to, at a minimum, preserve genetic linkages). Yet, more extreme management measures at the local scale are also needed. Such diverse viewpoints contribute to a fog of uncertainty in which managers must determine what information is necessary and sufficient to manage marine ecosystems.

One key unresolved issue is that there are no measurements of larval spillover downstream or near protected reefs. Spillover recorded to date applies only to adult fish (*2*–*4*). Although this illustrates the need for more targeted research on connectivity (*5*, *6*), it also makes it more difficult to argue for more and/or larger no-take reserves at stakeholder, manager, and governmental levels. Even if larval subsidies from MPAs are present, they may have little measurable effect on recruitment because of fewer recruitment habitats in unprotected or degraded reefs. Thus, marine reserves may be necessary but not sufficient for marine conservation (*7*).

Given our scientific uncertainty, managers should pursue other locally supported measures such as limiting fishing to specific sizes of fish, banning exports, or eliminating certain fishing methods. Our failure to effectively manage marine ecosystems may have less to do with gaps in our science than it does in getting buy-in from stakeholders. A diverse and locally adapted management toolbox may be more effective than creating more no-take areas where compliance will be low (*8*).

ROBERT S. STENECK,[1]* ROBERT K. COWEN,[2] CLAIRE B. PARIS,[2] ASHWANTH SRINIVASAN[3]

[1]School of Marine Sciences, Darling Marine Center, University of Maine, 193 Clarks Cove Road, Walpole, ME 04573, USA. [2]Division of Marine Biology and Fisheries, [3]Division of Meteorology and Physical Oceanography, Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Miami, FL, 33149, USA.

*To whom correspondence should be addressed. E-mail: steneck@maine.edu

References

1. A. Pérez-Ruzafa, M. González-Wangüemert, P. L'enfant, C. Marcos, J. A. García-Charton, *Biol. Conserv*. **129**, 244 (2006).
2. C. M. Roberts, J. A. Bohnsack, F. Gell, J. P. Hawkins, R. Goodridge, *Science* **294**, 1920 (2001).
3. G. R. Russ, B. Stockwell, A. C. Alcala, *Mar. Ecol. Prog. Ser*. **292**, 1 (2005).
4. G. R. Russ, A. C. Alcala, A. P. Maypa, *Mar. Ecol. Prog. Ser.* **264**, 15 (2003).
5. R. K. Cowen, G. Gawarkiewicz, J. Pineda, S. Thorrold, F. Werner, Report of a Workshop to Develop Science Recommendations for the National Science Foundation (National Science Foundation, Arlington, VA, 2003).
6. P. F. Sale *et al.*, *Trends Ecol. Evol.* **20**, 74 (2005).
7. G. Allison, J. Lubchenco, M. H. Carr, *Ecol. Applic.* **8**, S79 (1998).
8. S. C. Jameson, M. H. Tupper, J. M. Ridley, *Mar. Pollut. Bull.* **44**, 117 (2002).

TECHNICAL COMMENT ABSTRACTS

COMMENT ON "On the Regulation of Populations of Mammals, Birds, Fish, and Insects" IV

Elizabeth Peacock and David L. Garshelis

Sibly *et al.*'s (Reports, 22 July 2005, p. 607) contention that density dependence acts strongly on low-density animal populations irrespective of body size contradicts many long-term studies of large mammals. Their findings were distorted by harvest records, which may poorly reflect population trend. Omitting unreliable data, their massive data set is reduced to only one case for large mammals.

Full text at www.sciencemag.org/cgi/content/full/313/5783/45a

RESPONSE TO COMMENT ON "On the Regulation of Populations of Mammals, Birds, Fish, and Insects"

Richard M. Sibly, Daniel Barker, Michael C. Denham, Jim Hone, Mark Pagel

Our conclusions are unaffected by removal of the time series identified by Peacock and Garshelis as harvest data. The relationship between a population's growth rate and its size is generally concave in mammals, irrespective of their body sizes. However, our data set includes quality data for only five mammals larger than 20 kilograms, so strong conclusions cannot be made about these animals.

Full text at www.sciencemag.org/cgi/content/full/313/5783/45b

Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects" IV

Elizabeth Peacock[1]* and David L. Garshelis[2]

Sibly *et al.*'s (Reports, 22 July 2005, p. 607) contention that density dependence acts strongly on low-density animal populations irrespective of body size contradicts many long-term studies of large mammals. Their findings were distorted by harvest records, which may poorly reflect population trend. Omitting unreliable data, their massive data set is reduced to only one case for large mammals.

Sibly *et al*. (*1*) used a massive collection of data on wild vertebrates and invertebrates [Global Population Dynamics Database (GPDD)] to assess density-related changes in population growth rates (*pgr*). In the simple logistic population growth model, *pgr* declines linearly with density, under the assumption that each additional animal's resource use has the same negative effect on the survival and reproduction of others in the population. Sibly *et al*. found, however, that from insects to mammals, *pgr* versus population size generally showed a concave relationship, indicating that additional individuals have a more depressing effect at low density. This result, while not novel for short-lived animals, was contrary to expectation for large, long-lived ("*K*-selected") mammals, which were thought to exhibit weak density dependence until very near carrying capacity, *K* (i.e., a convex curve) (*2*, *3*). A chief implication is that overharvested populations of large mammals (well below *K*) may take much longer to recover than expected. We show that this contention is unfounded.

Many time series in the GPDD are harvest data, not population data. Harvests mirror population size only when hunter numbers or success rates directly respond to changes in animal abundance, as may be the case for steadily growing or declining populations, or predictably cyclic populations (*4*–*6*). Even so, harvests may overestimate variation in population size by several orders of magnitude (*7*) or may underestimate variation if hunter numbers are restricted by management agencies.

Harvest trends may reflect changing demand for the hunted species. Although this relationship is well known from several centuries of commercial harvests of furbearing mammals (*8*), Sibly *et al*.'s analysis included numerous fur harvest records (Fig. 1). Harvests also may vary with economic and social conditions (affecting numbers of hunters, their effort, and their motivation) (Fig. 2) and with hunting methods (efficiency). Wildlife managers are as likely to be alarmed that rising harvests are causing a population decline as they are to take comfort that such harvests are tracking an increasing population. Treatises on the subject of population monitoring do not even address the use of raw harvest numbers to assess population trend (*9*, *10*). Even adjusted for hunter effort, harvests may poorly reflect changes in population size because of other confounding variables (*11*).

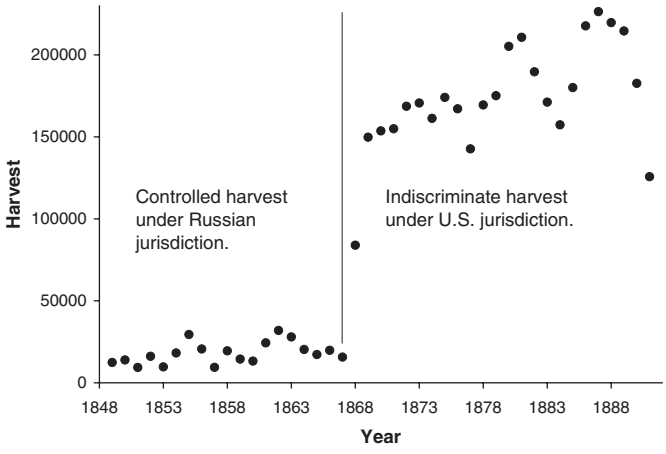


Fig. 1. Northern fur seal (*Callorhinus ursinus*) harvests in Alaska, considered by Sibly *et al*. to be reflective of population size and a case of a concave growth curve for a large mammal (GPDD number 3804, $\theta = -3.98$). The sudden increase in harvest in 1868 reflected the purchase of Alaska by the United States the previous year (indicated by line). During previous Russian ownership, low harvests were a result of a ban on killing female seals and limits on the killing of males, yielding high populations. These restrictions were lifted after the U.S. purchase, promoting more interest in sealing, including the commencement of pelagic sealing, leading to near extermination of the species (*21*). Hence, these harvest data reflect the inverse of population size. Long-term data on living populations of this species consistently show density effects only near *K* (*15*, *22*).

Explanatory information provided with the GPDD indicates that harvest data are regarded as "highly unreliable" as population data. Sibly *et al*. asserted that their conclusions remained unchanged when certain categories of unreliable data (including harvest data) were omitted; however, they did not indicate the taxonomic breadth or number of remaining cases. We scrutinized the 977 records of GPDD mammal data and found that 65% were harvest data. We eliminated these and other data regarded as low quality, as Sibly *et al*. had done, and then continued filtering data using their procedures (table S1). This left 14 time series where θ, the parameter describing curvature of the *pgr* function, was specified and within the ascribed confidence limits. Only three of these valid data sets were large mammals, only one of which exhibited density dependence (with a slightly convex *pgr* curve). Two had concave *pgr* functions, but in both of these, *pgr* related more to rainfall than to density (*12*, *13*); strong environmental variation needs to be accounted for before assessing density-dependent relationships (*14*).

Large, amalgamated databases like the GPDD may be a potential source for new insights; alternatively, as in this case, interpretations of second-hand data may lead to erroneous inferences. Careful, long-term studies of population change in large mammals, difficult as they are, have consistently concluded that these species exhibit strong density dependence only near *K*; thus, maximum net increase in population size occurs between *K* and *K*/*2* (*14*–*17*). That principle is inherent in national legislation governing exploitation of marine mammals (Marine Mammal Protection Act) (*18*) and is relevant in projecting population growth and recovery of other previously overharvested, *K*-selected species, such as grizzly bears (*19*).

Sibly *et al*.'s analysis provides little reason to challenge this paradigm of population growth for large mammals. Their regression, indicating that large mammals have a more concave θ than smaller mammals, is not meaningful given that θ was not obtained from reliable population data. Harvest data may yield the perception of a concave relationship even if *pgr* versus population size is convex. Accordingly, Sibly *et al*.'s

[1]Department of Biology, University of Nevada–Reno, Reno, NV 89557, USA. [2]Minnesota Department of Natural Resources, Grand Rapids, MN 55744, USA.

*To whom correspondence should be addressed. E-mail: epeacock@gov.nu.ca

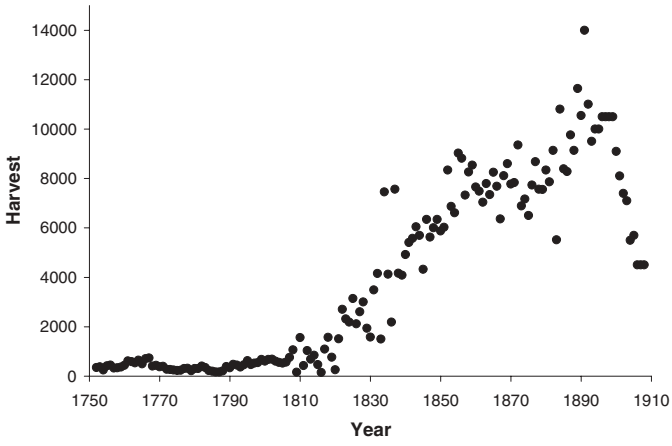


Fig. 2. Bear (*Ursus* spp.) harvests in North America, taken to be reflective of changes in population size. Sibly *et al*. considered this a case of a concave growth curve for a large mammal (GPDD number 3769, $\theta = -12.6$). Increasing bear harvests during this period reflected increased European settlement of western North America, and commensurate increased killing and widespread extirpation of bears (*23*). Hence, these harvest data reflect the inverse of population size, up to the point where bear numbers collapsed to the extent that kills began to decline.

warnings of "dangerous consequences" related to slower-than-expected recovery of large mammal populations seems unwarranted. Nevertheless, because large mammals have low *pgr*, managers should err on the side of caution and treat local circumstances on a case-by-case basis, because there are likely to be some situations where growth rates are depressed at low density (*20*).

References and Notes

1. R. M. Sibly, D. Barker, M. C. Denham, J. Hone, M. Pagel, *Science* **309**, 607 (2005).
2. C. W. Fowler, *Ecology* **62**, 602 (1981).
3. C. W. Fowler, *Curr. Mammal.* **1**, 401 (1987).
4. I. M. Cattadori, D. T. Haydon, S. J. Thirgood, P. J. Hudson, *Oikos* **100**, 439 (2003).
5. I. M. Cattadori, D. T. Haydon, P. J. Hudson, *Nature* **433**, 737 (2005).
6. N. C. Stenseth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15430 (1998).
7. X. Lambin, C. J. Krebs, R. Moss, N. C. Stenseth, N. G. Yoccoz, *Science* **286**, 2425a (1999).
8. M. Novak, J. A. Baker, M. E. Obbard, B. Malloch, Eds., *Wild Furbearer Management and Conservation in North America* (Ontario Ministry Natural Resources, Toronto, Ontario, Canada, 1987).
9. G. Caughley, *Analysis of Vertebrate Populations* (Wiley, Chichester, UK, 1977).
10. B. K. Williams, J. D. Nichols, M. J. Conroy, *Analysis and Management of Animal Populations* (Academic Press, San Diego, CA, 2002).
11. J. I. Schmidt, J. M. Ver Hoef, J. A. K. Maier, R. T. Bowyer, *J. Wildl. Manage.* **69**, 1112 (2005).
12. N. Owen-Smith, *J. Anim. Ecol.* **59**, 893 (1990).
13. J. P. Hanby, J. D. Bygott, C. Packer, in *Serengeti II: Dynamics, Management, and Conservation of an Ecosystem*, A. R. E. Sinclair, P. Arcese, Eds. (Univ. Chicago Press, Chicago, 1995), pp. 315–331.
14. N. Owen-Smith, *Ecol. Monogr.* **76**, 93 (2006).
15. C. W. Fowler, *Mar. Mammal. Sci.* **6**, 171 (1990).
16. S. Jeffries, H. Huber, J. Calambokidis, J. Laake, *J. Wildl. Manage.* **67**, 208 (2003).
17. A. R. E. Sinclair, *Phil. Trans. R. Soc. London Ser. B* **358**, 1729 (2003).
18. B. L. Taylor, P. R. Wade, D. P. DeMaster, J. Barlow, *Conserv. Biol.* **14**, 1243 (2000).
19. M. S. Boyce, B. M. Blanchard, R. R. Knight, C. Servheen, *Int. Assoc. Bear Res. Manage Monogr. Ser.* **4** (2001).
20. H. U. Wittmer, A. R. E. Sinclair, B. N. McLellan, *Oecologia* **144**, 257 (2005).
21. A. Y. Roppel, S. P. Davey, *J. Wildl. Manage.* **29**, 448 (1965).
22. M. A. Etnier, *Can. J. Fish. Aquat. Sci.* **61**, 1616 (2004).
23. D. J. Mattson, T. Merrill, *Conserv. Biol.* **16**, 1123 (2002).

24. We thank two anonymous reviewers for their suggestions.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/45a/DC1
Table S1

21 March 2006; accepted 25 April 2006
10.1126/science.1127705

Response to Comment on "On the Regulation of Populations of Mammals, Birds, Fish, and Insects"

Richard M. Sibly,[1]* Daniel Barker,[2] Michael C. Denham,[3] Jim Hone,[4] Mark Pagel[1]

Our conclusions are unaffected by removal of the time series identified by Peacock and Garshelis as harvest data. The relationship between a population's growth rate and its size is generally concave in mammals, irrespective of their body sizes. However, our data set includes quality data for only five mammals larger than 20 kilograms, so strong conclusions cannot be made about these animals.

Our analysis of the relationship between the abundance of a population and its rate of growth (*pgr*) used data from 4926 different populations in the Global Population Dynamics Database, GPDD (*1*). After filtering for data quality, there remained 1780 time series of mammals, birds, fish, and insects. We reported that the relationship between *pgr* and abundance is concave in 78% of species and that there is little variation between the major taxonomic groups. In their comment, Peacock and Garshelis (*2*) suggest that many of the mammal data sets in the GPDD are harvest data and are therefore uninformative about population size; that when unreliable large mammal data are removed, only one case remains; and that the general conclusions do not, therefore, apply to large mammals. We do not accept that all harvest data should be removed from the GPDD, because several analyses have reported that, where checks could be made, harvest sizes were proportional to population sizes (*3*–*6*). Nevertheless we here rerun the mammal analyses removing the data sets that Peacock and Garshelis identify as harvest data and show that our original conclusions are unaffected by this reanalysis.

The GPDD was set up and is used as a collection of time series of population abundance (*7*–*12*). These measures of population abundance come from a wide range of sources and do vary considerably in data quality, as discussed on the GPDD Web site (*13*). In the absence of accurate, objective measures of data quality, the constructors of the GPDD used subjective judgment to assign a rank for apparent data quality on a scale from 1 (low) to 5 (high), using criteria such as the type of environment sampled, the species in question, the area of the sampling site, and the method of sampling. The GPDD does not classify the data as being harvest or otherwise. The two time series highlighted in the figures in (*2*) are graded 1. We accept that they should not have been included in our original analysis.

Our approach to the issue of variation in data quality was to run the analyses both on the whole data set and on a data set restricted to the top two GPDD gradings (i.e., grades 4 and 5). As we reported, we found no difference in conclusions between these two analyses, and this gave us confidence in the general validity of our results (*1*). In the case of the mammals, there are 19 time series, corresponding to 14 species, in the top two GPDD gradings.

We repeated our analysis without the harvest data identified by Peacock and Garshelis but now including the top four GPDD gradings. Peacock and Garshelis identified 52 mammal time series in this category, and these correspond to 29 separate species [see table S1 in (*2*)]. Of these species, 23 have a concave relationship between *pgr* and abundance, and 6 are convex (Fig. 1A, chi-square test of equal proportions: $\chi_1^2 = 10.0$, $P < 0.005$). The proportion that are concave is 0.79, similar to the overall figure of 0.78 that we reported for birds, fish, mammals, and insects (*1*). There is no evidence of an increase in θ with body size; indeed, the reverse is the case (Fig. 1A, Spearman's $\rho_{22} = -0.37$, $P = 0.07$). If harvest species are included in the above analysis, the proportion of curves that are concave is 0.80, and again θ decreases rather than increases with body size (Fig. 1B, Spearman's $\rho_{32} = -0.40$, $P = 0.02$).

Restricting attention to the top two GPDD gradings, four of the five mammal species with body mass >20 kg have concave relationships (θ < 1). These are *Canis lupus, Panthera leo, Phoca vitulina*, and *Tragelaphus strepsiceros*. Only *Mirounga leonina* has θ > 1. Peacock and Garshelis discarded all four species with concave relationships. They discarded two on the grounds that the esti-

[1]School of Biological Sciences, University of Reading, Whiteknights, Reading RG6 6AJ, UK. [2]Sir Harold Mitchell Building, School of Biology, University of St. Andrews, St. Andrews, Fife, KY16 9TH, UK. [3]Statistical Sciences Europe, GlaxoSmithKline Research and Development Limited, New Frontiers Science Park, Third Avenue, Harlow, Essex CM19 5AW, UK. [4]Institute for Applied Ecology, University of Canberra ACT 2601, Australia.

*To whom correspondence should be addressed. E-mail: r.m.sibly@reading.ac.uk

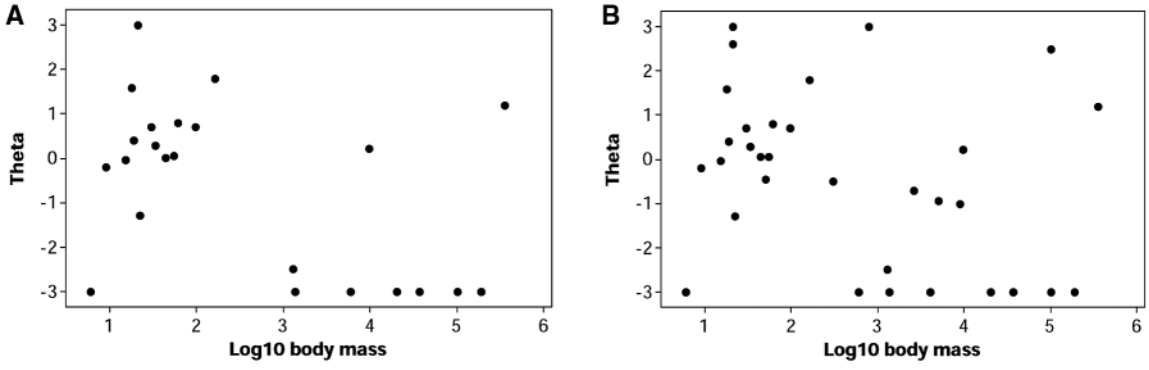


Fig. 1. Plots of θ against body mass *g* for mammals, one point per species. θ is the parameter fitted by (*1*) that specifies the curvature of the relationship between *pgr* and abundance. θ < 1 indicates a concave relationship, θ > 1 convex. θ values above 3 have been plotted as 3 (and similarly for θ < –3) to avoid giving undue prominence to large values of θ. Body mass data are not available for all species, but in some cases it has been possible to supplement GPDD data from (*16*). (**A**) The GPDD higher quality data (grades 2 to 5) classified by Peacock and Garshelis as nonharvest data. (**B**) As (A), but including harvest data. The GPDD data ID reference numbers of the time series from which these data were calculated are (A) 63, 64, 3605, 3606, 3607, 1318, 10532, 10543, 65, 66, 1301, 3512, 3514, 3581, 1308, 3969, 3970, 2723, 2724, 3888, 3515, 1295, 1296, 10545, 3580, 3256, 1328, 1330, 1331, 1332, 1333, 3404, 3198, 1273, 1274, 1275, 1276, 10533, 3466, 1327, 3736, 3582, 10541, 1292, 10535, 10540, 3603, 3268, 3877, 3879, 3880, and 3881. (B) also includes 1343, 1345, 1540, 1541, 3569, 3655, 3702, 3705, 3706, 3708, 3709, 3721, 3724, 3731, 4006, 4007, 10538, 10539, 11012, and 11015.

mates of θ and its lower confidence limit appeared the same, i.e., equal to –31.6 in the GPDD. In reality, however, $–31.6 = –10^{-1.5}$ is the lowest value we could reliably estimate, so our estimates of θ are never below –31.6, and where we report that a lower confidence limit is –31.6, this means ≤ –31.6, as specified in our supporting online material for (*1*). Peacock and Garshelis argue that the other two species should be discounted because an extrinsic factor (rainfall) affected *pgr* independent of abundance. Extrinsic variation is, however, universal, and it would be difficult to estimate the *pgr*-abundance relationship without it.

Finally, Peacock and Garshelis imply the existence of a number of careful long-term studies showing the existence of convex *pgr*-abundance relationships, but of the four references they cite, two do not estimate *pgr* and one is based on unpublished data (*14*). We agree that the fourth, just published, is an exemplary study (*15*).

In summary, our original conclusions are unaffected by removal of the data sets identified by Peacock and Garshelis as harvest data. In particular, our conclusion is correct that, for the data in the GPDD, the relationship between a population's growth rate and its size is generally concave in mammals. Four of the five mammal species with body mass >20 kg have concave relationships, but strong conclusions cannot be drawn from just five cases, and more studies of large mammals are needed.

References

1. R. M. Sibly, D. Barker, M. C. Denham, J. Hone, M. Pagel, *Science* **309**, 607 (2005).
2. E. Peacock, D. L. Garshelis, *Science* **313**, 45 (2006); www.sciencemag.org/cgi/content/full/313/5783/45a.
3. N. C. Stenseth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15430 (1998).
4. I. M. Cattadori, D. T. Haydon, P. J. Hudson, *Nature* **433**, 737 (2005).
5. M. C. Forchhammer, T. Asferg, *Proc. R. Soc. London B. Biol. Sci.* **267**, 779 (2000).
6. I. M. Cattadori, D. T. Haydon, S. J. Thirgood, P. J. Hudson, *Oikos* **100**, 439 (2003).
7. J. M. Halley, K. I. Stergiou, *Fish Fish.* **6**, 266 (2005).
8. D. H. Reed, G. R. Hobbs, *Anim. Conserv.* **7**, 1 (2004).
9. D. H. Reed, J. J. O'Grady, J. D. Ballou, R. Frankham, *Anim. Conserv.* **6**, 109 (2003).
10. P. Inchausti, J. Halley, *Science* **293**, 655 (2001).
11. W. F. Fagan, E. Meir, J. Prendergast, A. Folarin, P. Karieva, *Ecol. Lett.* **4**, 132 (2001).
12. B. E. Kendall, J. Prendergast, O. N. Bjornstad, *Ecol. Lett.* **1**, 160 (1998).
13. Global Population Dynamics Database, http://cpbnts1.bio.ic.ac.uk/gpdd/quality.htm.
14. A. R. E. Sinclair, *Philos. Trans. R. Soc. London Ser. B* **358**, 1729 (2003).
15. N. Owen-Smith, *Ecol. Monogr.* **76**, 93 (2006).
16. S. K. M. Ernest, *Ecology* **84**, 3402 (2003).

21 February 2006; accepted 6 June 2006
10.1126/science.1124973

HISTORY OF SCIENCE

Galileo, Scientific Entrepreneur

Paula E. Findlen

The life and work of the early 17th-century mathematician and astronomer Galileo Galilei have been presented in many different guises, depending on which aspects of his highly publicized and controversial career authors have chosen to explore. Historians fascinated with his work as an observational astronomer have examined Galileo as an instrument-maker, discoverer, and inaugurator of a new approach to understanding the heavens. Historians interested in his contributions to physics and mechanics have explored the long evolution of his ideas and experiments culminating in the appearance of his *Two New Sciences* (1638). Scholars from many disciplines have also found Galileo's work to be a fascinating example of the interconnections between the arts and sciences, since he drew liberally upon Renaissance literary and artistic traditions to construct a more accessible form of knowledge. Most famously, however, we have studied Galileo because of his 1633 trial and condemnation by the Roman Catholic Church for advocating heliocentrism. This crucial episode has cast a long shadow on the often vexed relations between science and faith, at times obscuring the many other reasons why Galileo is worth studying.

Mario Biagioli's work has played an important role in the revival of Galileo studies since the late 1980s. In an earlier book (*1*), Biagioli discussed the process by which a Paduan mathematics professor parlayed his discovery in 1609 of four moons of Jupiter into an appointment by 1610 as court mathematician and philosopher to the Grand Duke of Tuscany. That book offered a fascinating meditation on the relationship between science and power and explored the consequences of doing science at court in the shifting political and religious climate of early 17th-century Italy. The Galileo who emerged in its analysis was a consummate political actor who understood what society might offer the scientist and vice versa. He responded to the challenge of making knowledge not just for a scholarly community but for everyone to debate and understand.

Galileo's Instruments of Credit focuses on the period spanning the publication of *The Sidereal Messenger* (1610), which announced Galileo's first telescopic discoveries, to the condemnation of heliocentrism by the Catholic Church (1616). Biagioli's goal is to explain how Galileo gained credit during this critical period in the formation of his science: "As Galileo was constructing new instruments and claims, he was also constructing the economy in which his claims could be credited." Biagioli explores the wide variety of strategies used by Galileo to achieve these goals and discusses the responses they generated.

In many respects, the book might equally well be titled "Galileo, Scientific Entrepreneur." Biagioli is primarily concerned with Galileo as an actor in a marketplace of inventions and ideas, competing with other astronomers, natural philosophers, and inventors for rewards and favors and trying to persuade patrons and readers of the reliability of his observations and interpretations. Successive chapters explore the process by which this occurred, beginning with the avenues Galileo used to persuade the Medici and others that his first telescopic observations were reliable. Biagioli then discusses the more difficult presentation of competing illustrations and interpretations of sunspots by Galileo and his Jesuit rival Christoph Scheiner in 1612–1613. Biagioli's account culminates in a consideration of the strategies Galileo adopted, in his unpublished "Letter to the Grand Duchess Christina" (1615), to convince readers of the equivalency between the books of nature and Scripture.

Galileo's Instruments of Credit
Telescopes, Images, Secrecy
by Mario Biagioli
University of Chicago Press, Chicago, 2006. 312 pp. $35, £22.50. ISBN 0-226-04561-7.

The first economy that Biagioli explores, which coincided with Galileo's move to Florence from Padua, is one of invention and print. How did Galileo present his telescope to others? Arguing that Galileo alternated "between secrecy and transparency," Biagioli discusses how Galileo limited access to his telescopes, reinforcing his role as the primary observer, while presenting his findings in cinematic form so that readers would be persuaded through words and images that planets and satellites moved in the way that he described. Galileo sought to establish a monopoly of invention that fueled his claims to discovery. The concept of "pictorial narrative" played an even greater role in his defense of sunspots as proof that the Sun rotated on its axis against Scheiner's alternative explanations. How might competing depictions of nature persuade readers of their conflicting philosophical claims?

Effective instrument. Using this telescope, Galileo discovered four moons of Jupiter.

The second major economy, reflecting Galileo's growing sense of Rome as a critical audience for his observations and theories, concerns truth and authority. In his final chapter, Biagioli explores the circumstances that led Galileo to present nature as a "book." What kind of book was nature in comparison with Scripture? Who should read and interpret it? Arguing that the true recipient of his 1615 letter was Cardinal Robert Bellarmine rather than the Grand Duchess Christina of Lorraine, Biagioli offers a close reading of this important text. He examines its failure to persuade theologians that others should be qualified to interpret this aspect of God's creation for them, while also analyzing Galileo's rhetorical strategies that might account for the seemingly serendipitous emergence of the scientific interpreter of nature: "Galileo fashioned himself as the reader whom God had not planned to exist, but whose existence he had not explicitly forbidden either."

Galileo is the centerpiece of the book, but Biagioli has not written a study focused solely on him. Almost 30 pages in the chapter on the construction of scientific authority as an "investment process" discuss the early Royal Society as a corporate counterpart to Galileo's highly personal efforts to manage the problems of secrecy, distance, and authority. The account of Galileo's image-making also draws on other examples of scientific illustrations from the 17th century. While this material enriches Biagioli's presentation, I am less confident about the broader methodological premise of the book. Although there is plenty of good

The reviewer is at the Department of History, Stanford University, Stanford, CA 94305–2024, USA. E-mail: pfindlen@stanford.edu

Published by AAAS

specific information about Galileo in this study, Biagioli prefers to examine Galileo less as a historical actor and more as an object of meditation for contemporary science studies, certain aspects of literary theory [namely, in his discussion of the "Letter to the Grand Duchess Christina," Jacques Derrida (*2*)], and attendant theoretical reflections on such subjects as networks, distance, blackboxing (simply citing references in support of a position), inscription, and supplementation. The author's goal, I suspect, is to debate key aspects of the theoretical formulation of science studies with material drawn from the Galileo archive.

Throughout the book, Biagioli constantly compares Galileo's activities with modern-day examples of scientific practice and entrepreneurship, drawing liberally from his readings in contemporary science studies and debates about intellectual property. This approach surely has its virtues, but readers not immersed in either of these literatures will at times be puzzled, perhaps even put off, by the construction of the author's arguments. By theorizing Galileo so explicitly, Biagioli has made him of greater interest to some readers. But I suspect the author has also sacrificed the power of a historical narrative, which he could have written by digging more deeply into the documents of Galileo's own world in order to give us a richer and more context-specific account of the nature of invention and entrepreneurship at the end of the Renaissance.

Galileo's Instruments of Credit is worth reading, both to see where studies of Galileo are going and to understand the uneasy but always interesting relationship between the history of science and contemporary science studies. Although lacking the unifying thesis and narrative power of the author's earlier book (*1*), it nonetheless offers a number of interesting insights into how to understand the uses of instruments and images in early modern science. I hope Biagioli will pursue in greater detail the historical dimensions of the issues he has raised.

References

1. M. Biagioli, *Galileo, Courtier: The Practice of Science in the Culture of Absolutism* (Univ. Chicago Press, Chicago, 1993).
2. J. Derrida, *Of Grammatology* G. C. Spivak, Transl. (Johns Hopkins Univ. Press, Baltimore, MD, 1976).

10.1126/science.1129563

EXHIBITS: HISTORY OF SCIENCE

Islamic Inventiveness

John Pickstone

Most universities include staff and students from many ethnic backgrounds; they often know something of the contributions from their own cultural traditions to the common heritage of science and technology. Sometimes they interact with professional historians of science, and a guest lecture or two is the usual result. But one such interaction, 30 years ago, grew into the lively exhibition *1001 Inventions: Discover the Muslim Heritage in Our World* and its attractive accompanying book.

1001 Inventions
Discover the Muslim Heritage in Our World

Salim Al-Hassani, originator

Museum of Science and Industry, Manchester, through 3 September 2006. www.1001inventions.com

1001 Inventions
Muslim Heritage in Our World

Salim Al-Hassani, Ed.

Foundation for Science, Technology and Civilisation, Manchester, UK, 2006. 372 pp. £25. ISBN 0-9552426-0-6.

At the Institute of Science and Technology (UMIST), which is now part of the University of Manchester, the then principal, Lord Vivian Bowden, became interested in the history of the Muslim empire, especially its financial arrangements. He arranged a collaboration between historians of the Middle East, historians of science, and some interested scientists. That faded away, but one of the participants, Salim Al-Hassani (a professor of mechanical engineering at UMIST), realized how little most people knew about the Muslim contributions to science and technology. Encouraged by Donald Cardwell (UMIST's former historian of science and technology and the founder of Manchester's Museum of Science and Industry), he collected like-minded colleagues and, with support from specialist historians, developed a Web site (www.muslimheritage.com).

After September 2001, the project gained new momentum, as the media promoted histories that would reduce the apparent cultural difference between Western and Islamic civilizations. Thus the current exhibition in Manchester can draw on some very professional videos, as well as models, attractive posters, and a few primary exhibits. These are organized around seven spaces, from home and school through market, hospital, and town to the world and universe.

When Britain learned with horror that the July 2005 bombings in London were planned and executed by Asians educated in the United Kingdom, the political import of the exhibition's theme increased further. Young whites should know about the many cultures on which the Western tradition draws; young Asians should recognize the depth and breadth of their historical tradition, including its science.

The exhibition's message is simple: see how much of the common heritage of science, technology, and medicine was pioneered in the Islamic world between 500 and 1500—astrolabes, algebra, and alchemy; coffee, clocks, and carpets; eye surgery, setting fractures, and hospitals; universities and professorial chairs. The audience will certainly get the point, and one young visitor was left wondering "Is there anything Muslims did not discover?"

Of course, such messages are a preliminary to real history. We would all benefit from knowing under what conditions our many and various traditions have proved fertile or infertile for invention and how those traditions have interacted over time. To that end, we need global histories of science that deal adequately with social contexts. But that is a longer term goal. For now, we can be grateful to Al-Hassani for his devotion, and we can hope that the exhibition and its products will gain wide circulation around the world.

10.1126/science.1129062

Working replica. This water clock is described by Al-Jazari in *The Book of Knowledge of Ingenious Mechanical Devices* (1206).

The reviewer is at the Centre for the History of Science, Technology and Medicine, University of Manchester, Oxford Road, Manchester M13 9PL, UK. E-mail: john.pickstone@manchester.ac.uk

BIODIVERSITY

Confronting Amphibian Declines and Extinctions

Stopping further global losses of amphibian populations and species requires an unprecedented conservation response.

Joseph R. Mendelson III,* Karen R. Lips, Ronald W. Gagliardo, George B. Rabb, James P. Collins, James E. Diffendorfer, Peter Daszak, Roberto Ibáñez D., Kevin C. Zippel, Dwight P. Lawson, Kevin M. Wright, Simon N. Stuart, Claude Gascon, Hélio R. da Silva, Patricia A. Burrowes, Rafael L. Joglar, Enrique La Marca, Stefan Lötters, Louis H. du Preez, Ché Weldon, Alex Hyatt, José Vicente Rodriguez-Mahecha, Susan Hunt, Helen Robertson, Brad Lock, Christopher J. Raxworthy, Darrel R. Frost, Robert C. Lacy, Ross A. Alford, Jonathan A. Campbell, Gabriela Parra-Olea, Federico Bolaños, José Joaquin Calvo Domingo, Tim Halliday, James B. Murphy, Marvalee H. Wake, Luis A. Coloma, Sergius L. Kuzmin, Mark Stanley Price, Kim M. Howell, Michael Lau, Rohan Pethiyagoda, Michelle Boone, Michael J. Lannoo, Andrew R. Blaustein, Andy Dobson, Richard A. Griffiths, Martha L. Crump, David B. Wake, Edmund D. Brodie Jr.

The Panamanian Golden Frog, *Atelopus zeteki*, is nearly extinct in the wild as a combined result of habitat change, illegal collecting, and fungal disease; the species is currently secure in several ex situ programs.

CREDIT: RON HOLT

Amphibian declines and extinctions are global and rapid: 32.5% of 5743 described species are threatened, with at least 9, and perhaps 122, becoming extinct since 1980 (*1*). Species have disappeared across the entire taxonomic group and in nearly all regions of the planet. These figures are probably underestimates as entire clades of species are threatened. For example, of the 113 species of harlequin toads (genus *Atelopus*), 30 are possibly extinct, and only 10 have stable populations (*2*). Nearly a quarter of known amphibian species were deemed "data-deficient" with respect to conservation status in the recent global assessment (*1*). Losing biodiversity at this taxonomic scale impacts ecosystem goods and services [e.g. (*3*, *4*)]. As amphibian species disappear, we also lose their untapped potential for advances in biomedicine and biotechnology in general (*5*).

Losses result from familiar threats (land-use change, commercial overexploitation, and exotic species) and from the emerging infectious disease chytridiomycosis, caused by the fungus *Batrachochytrium dendrobatidis* (*Bd*). Predictions are that within 4 to 6 months of *Bd* arrival at a site where it has not previously been present, ~50% of amphibian species and ~80% of individuals may disappear (*6*). Global climate change may be encouraging local conditions ideal for *Bd*'s persistence and/or spread (*7*), commercial trade of wildlife may also contribute (*8*), and pollution may increase susceptibility of species to pathogens (*9*, *10*). Traditional programs and current laws and policies alone are insufficient to address global threats that cross boundaries of reserves and nations.

Global leaders in research, conservation, and policy agreed on an Amphibian Conservation Action Plan (ACAP) and Declaration in 2005 (see Supporting Online Material). A new, international body was recommended to coordinate and facilitate conservation programs for amphibians and to garner and administer funds. Thus, we call for formation of The Amphibian Survival Alliance (ASA)—led by an international secretariat of the Amphibian Specialist Group of the Species Survival Commission of IUCN (World Conservation Union). An initial 5-year budget requires at least U.S.$400 million.

Conservation activities should remain in affected countries where possible, with coordination and support through ASA, to engage and employ local scientists. A special initiative would be regional centers for disease research and captive breeding. Centers would exist within government agencies, zoos, or universities and would be staffed by local scientists, wildlife managers, and conservationists. ASA would create and support readily available databases from the global network of centers, as well as research and training in countries with few amphibian experts. Such dedicated research capacity in affected regions is required for this global crisis, as well as to keep amphibian research and conservation at the forefront of policy-making.

Chytridiomycosis deserves especial attention because of its massive impacts on amphibian diversity (*11*, *12*). Natural-agent control of *Bd* or selecting for resistance in amphibians may be possible (*13*, *14*). In the meantime, we must implement coordinated in situ actions (e.g., surveys, monitoring, and habitat protection) and ex situ husbandry programs (e.g., survival-assurance and research colonies) at unprecedented scales. Amphibian salvage operations are possible at an ecosystem level (*15*). Ex situ programs may be the only option to avoid extinction for many species [e.g., Kihansi Spray Toad and Panamanian Golden Frog (*16*, *17*)], while research progresses on disease control, treatment, and evolution of resistance (*18*).

The ASA model builds on programs such as the Turtle Survival Alliance (*19*), Global Environment Facility (GEF) Coral Program (*20*), and an Australian threat abatement plan (*21*). Success will depend on a paradigm shift in the scale of the coordinated response, with stakeholders from the academic, conservation, zoo, ethics, policy, global change, private sector, and international biodiversity convention communities uniting for one goal. Support from individuals, governments, foundations, and the wider conservation community is essential.

*To whom correspondence should be addressed. E-mail: jmendelson@zooatlanta.org. All author affiliations can be found in the Supporting Online Material.

References and Notes

1. S. N. Stuart *et al., Science* **306**, 1783 (2004).
2. E. La Marca *et al., Biotropica* **37**, 190 (2005).
3. J. A. Pounds *et al.,Nature* **398**, 611 (1999).
4. M. R. Whiles *et al., Front. Ecol.* **4,** 27 (2006).
5. S. E. VanCompernolle *et al., J. Virol.* **79,** 11598 (2005).
6. K. R. Lips *et al., Proc. Natl. Acad. Sci. U.S.A.* **103**, 3165 (2006).
7. J. A. Pounds *et al., Nature* **439**, 161 (2006).
8. C. Weldon *et al., Emerg. Infect. Dis*. **10**, 2100 (2004).
9. J. Kiesecker, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9900 (2002).
10. T. B. Hayes *et al., Environ. Health Perspect.* **114** (suppl. 1), 40 (2006).
11. L. Berger *et al., Proc. Natl. Acad. Sci. U.S.A.* **95**, 9031 (1998).
12. P. Daszak *et al.*, *Science* **287**, 443 (2000).
13. R. N. Harris *et al., Ecohealth* **3**, 53 (2006).
14. R.W. R. Retallick *et al.*, *PLoS* **2,** 1 (2004).
15. J. R. Mendelson III, G. B. Rabb, *WAZA Proceedings*, World Association of Zoos and Aquariums, 60th Annual Meeting, New York, 2 to 6 October 2005, in press.
16. K. Krajick, *Science* **311,** 1230 (2006).
17. K. C. Zippel, *Herpetol. Rev*. **33**, 11 (2002).
18. H. McCallum, *Conserv. Biol.* **19**, 1421 (2005).
19. Turtle Survival Alliance (www.turtlesurvival.org).
20. Coral Reef Targeted Research and Capacity Building for Management (CRTR) (www.gefcoral.org)
21. Threat abatement plans, infection of amphibians with chytrid fungus resulting in chytridiomycosis (www.deh.gov.au/biodiversity/threatened/publications/tap/chytrid/)

Supporting Online Material
www.sciencemag.org/cgi/content/full/313/5783/48/DC1

10.1126/science.1128396

MATHEMATICS

Exploring Musical Space

Julian Hook

New mathematical approaches can elucidate abstract musical spaces and help our understanding of harmonic processes at work in musical compositions.

A musical score is a graph whose vertical axis represents pitch and whose horizontal axis represents time. Given this apparent simplicity, and the recognition since the time of Pythagoras that mathematical principles underlie many musical phenomena, it is perhaps surprising that our understanding of the mathematical structure of the spaces in which musical phenomena operate remains fragmentary. But as Tymoczko shows on page 72 (*1*), even the pitch domain is vastly more complex than it may first appear.

Enhanced online at www.sciencemag.org/cgi/content/full/313/5783/49

Simply deciding what pitch structures are appropriate and when two of them should be considered "the same" can be difficult. Are notes sharing the same name but separated by octaves, such as two different C's, to be regarded as the same? What about notes that differ only enharmonically, such as C-sharp and D-flat? (These are a single note on the piano but often serve two very different musical purposes.) Harmonic structures, moreover, are not single notes but chords. If two of the voices forming a chord exchange notes, is the resulting chord the same? Should a four-note chord with two C's be reduced to a three-note chord with one C? Does the tuning of the instrument matter?

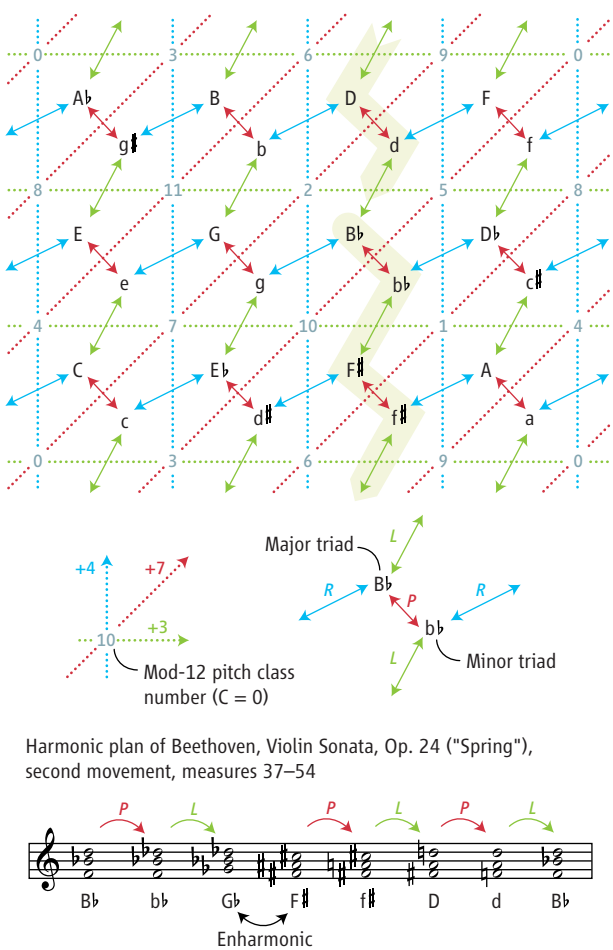
Musical sensibilities dictate different answers to these questions depending on the musical context and analytical objectives. Music theorists classify chords in categories such as major triads, grouping chords together with their translations in some appropriate space (in musical parlance, transposition), or in some cases also with their reflections (inversion). Western musicians are accustomed to a discrete view of pitch space, corresponding to the chromatic scale playable on the piano, but the general problem requires consideration of a larger space in which continuous pitch variation is possible. The various equivalence relations give rise to an assortment of quotient spaces (obtained by "gluing together" the points considered equivalent) and group structures acting on them. Tymoczko's work presented here is part of an ambitious project (*2*) that characterizes these spaces in great generality and relates the geometry of the spaces to the musical behavior of the chords that inhabit them.

The author is at the Jacobs School of Music, Indiana University, Bloomington, IN 47405, USA. E-mail: juhook@indiana.edu

Mathematical music theory, although terra incognita to practicing musicians and even to many professional music theorists, has in recent years blossomed into a sizable and multifaceted industry. Pitch-class set theory (*3*), the study of a discrete 12-note quotient space, was developed as a means of confronting the analytical challenges posed by "post-tonal" music of the 20th century, whose harmonic materials are more varied and complex than those in most earlier music. Diatonic set theory (*4*, *5*) investigates the subtle and beautiful relationship between the 12-note chromatic scale and diatonic scales such as the C major scale, with seven unequally spaced notes per octave (a scale type of great importance in many styles of music). Scale theory (*6*, *7*) studies structural properties of scales and their subscales more broadly, allowing variation in both chromatic and diatonic cardinalities and occasionally engaging considerations of tuning and acoustics.

In the past two decades, transformation groups acting on musical spaces have proven to be enormously fruitful models in a variety of settings. Transformations are mathematical functions that describe relationships between chords (or other musical entities); they often form algebraic groups and bear intimate relation to musical notions of interval (*8*).

A particularly active area is neo-Riemannian theory, which synthesizes modern group-theoretic techniques with inspiration drawn from the work of the prolific German musicologist Hugo Riemann (1849–1919) and his contemporaries. In its basic form (*9*, *10*), neo-Riemannian theory investigates certain transformational relationships among the 12 major and 12 minor triads in ways that are algebraically elegant, musically suggestive, and readily visualized in various forms of a graph known as a Tonnetz (tone network), in which the harmonic path traced by a musical composition may be plotted (see the figure).



Mapping music. A harmonic path in the neo-Riemannian Tonnetz.

In this representation, the mod-12 numbers in the background graph (dotted lines) designate pitch classes (pitches under the assumption of octave equivalence), from C = 0 through B = 11. They are arranged by musically important intervals: by perfect fifths (interval 7) diagonally (red), major thirds (interval 4) vertically (blue), and minor thirds (interval 3) horizontally (green). This two-dimensional graph is an unwrapped torus whose right edge is identified with the left and the top with the bottom. The graph in the foreground (solid arrows) depicts major and minor triads (labeled in upper and lower case, respectively), each positioned within the triangle formed by the corresponding pitch classes in the background; for example, the B-flat major triad comprises the notes B-flat (pitch class 10), D (pitch class 2), and F (pitch class 5). Each triad shares two of its notes with three different triads of the opposite mode, to which it is related by the transformations P (parallel, red arrows), R (relative, blue), and L [Leitton-

Published by AAAS

wechsel (leading tone exchange), green], which generate a dihedral group of order 24, isomorphic to the group of rotations and reflections of a 12-sided polygon. Each transformation exhibits efficient voice leading, preserving two pitch classes and moving the third by a small interval; arrows always cross dotted lines of the same color. The Beethoven progression (highlighted in the graph and expanded at the bottom of the figure) unfolds a PL-cycle that circumnavigates the torus, starting and ending in B-flat major, and illustrates that the composite transformation PLPLPL is the identity element of the group.

The Tonnetz is only one of many possible geometric representations of musical spaces (*11*), and recent studies have extended neo-Riemannian methods to larger and more powerful transformation groups, to other chord types besides triads, and in various other directions (*12*). In addition to group theory and other algebraic techniques, ideas from graph theory, combinatorics, geometry, and topology have found musical application. The work of Tymoczko *et al.* embraces all of these strategies in an innovative and wide-ranging investigation of chordal space. One of the great attractions of this work is its generality: It aims to describe what is in effect a "space of all chords" wherein the Tonnetz and many other familiar depictions of musical relationships appear as subspaces, projections, and cross sections. The spaces appearing here are of a type known as orbifolds, as they possess singularities—points where the geometry is not locally Euclidean. (The appeal to the recent topological concept of orbifolds is notable in a field that relies mainly on mathematics of a more classical vintage.) Other valuable contributions include a fresh perspective on the elusive notions of consonance and dissonance, connections between symmetries of the spaces and various musical practices, and many implications for the efficient chord-to-chord voice leading that has long been considered a hallmark of successful composition.

References

1. D. Tymoczko, *Science* **313**, 72 (2006).
2. C. Callender, I. Quinn, D. Tymoczko, paper presented at the John Clough Memorial Conference, University of Chicago, 7 to 10 July 2005.
3. A. Forte, *The Structure of Atonal Music* (Yale Univ. Press, New Haven, CT, 1973).
4. J. Clough, G. Myerson, *J. Music Theory* **29**, 249 (1985).
5. J. Clough, J. Douthett, *J. Music Theory* **35**, 93 (1991).
6. N. Carey, D. Clampitt, *Music Theory Spectrum* **11**, 187 (1989).
7. J. Clough, N. Engebretsen, J. Kochavi, *Music Theory Spectrum* **21**, 74 (1999).
8. D. Lewin, *Generalized Musical Intervals and Transformations* (Yale Univ. Press, New Haven, CT, 1987).
9. R. Cohn, *J. Music Theory* **41**, 1 (1997).
10. R. Cohn, *J. Music Theory* **42**, 167 (1998).
11. J. Douthett, P. Steinbach, *J. Music Theory* **42**, 241 (1998).
12. J. Hook, *J. Music Theory* **46**, 57 (2002).

10.1126/science.1129300

DEVELOPMENTAL BIOLOGY

Morphing into Shape

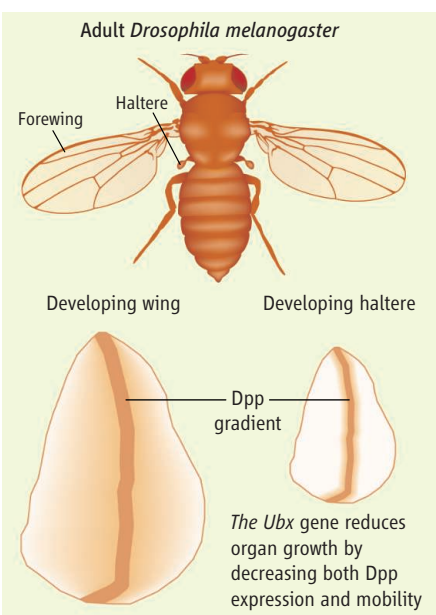
Two appendages of the fly, the haltere and the wing, grow to very different sizes. Limited expression and mobility of a growth morphogen is partly responsible for this difference.

David L. Stern

In 1917, British polymath D'Arcy Thompson proposed that the shapes of different organisms—say, a human and a chimpanzee—could be imagined as simple alterations of the same underlying pattern (*1*). Thompson famously demonstrated this idea by overlaying transformed Cartesian coordinates on drawings of related animals. This holistic view of organism shape inspired the British biologist Julian Huxley to point out that changes in shape can be thought of most simply as differences in the relative sizes of different body parts, thus reducing shape change to a more manageable problem (*2*). On page 63 of this issue, Crickmore and Mann (*3*) present a detailed analysis of the mechanisms controlling one striking difference in the relative size of two organs and uncover what may be a general mechanism of shape evolution.

In segmented organisms, such as flies and humans, similar structures that differ mainly in size and shape are produced in several locations along the main body axis. For example, humans produce arms and legs, largely using many of the same developmental mechanisms to pattern both organs. In fruit flies, two flying appendages, the wings and halteres (see the figure), also are built largely by shared developmental mechanisms. Halteres are delicate club-shaped organs that work like gyroscopes during flight. They evolved about 225 million years ago from more traditional-looking wings—such as the hind wings of butterflies—and have undergone a drastic reduction in size.

All of the differences between the wing and the haltere are determined by expression of a single "selector" gene called *Ultrabithorax* (*Ubx*), which is expressed in all cells of the developing haltere. When *Ubx* is experimentally removed from these cells, a fully formed wing grows instead of a haltere (*4*), revealing some of the underlying similarities between the two flight organs. *Ubx* somehow instructs other genes to alter the growth and development of haltere cells. In 1998, Weatherbee *et al.* (*5*) showed that *Ubx* regulates a battery of genes in the haltere, but until now we have not known precisely which genes are regulated to cause the greatest difference between the wing and the haltere: their fivefold difference in cell number in the adult.



How is organ size controlled? The *Ubx* gene is expressed in haltere cells, restricting the growth effect of the morphogen Dpp during development.

Crickmore and Mann focused their attention on how *Ubx* influences the activity of *decapentaplegic* (*dpp*), a gene that is one of the key regulators of wing growth. Dpp protein is produced by cells that lie in a line that is several cells wide along the middle of both the wing and the haltere. The protein is then secreted from these cells and diffuses to neighboring cells. When the Dpp protein binds to its receptor, Thickveins (Tkv), two things happen. First, a signal is triggered within the cell and this signal is interpreted as "grow more." Second, the Dpp protein is captured by the cell and eventually destroyed. Thus, Dpp protein diffuses away from the central cells and forms a gradient whose extent and steepness is controlled, at least in part, by the receptor Tkv.

Crickmore and Mann first noted that the width of the stripe of cells producing Dpp was narrower in the haltere than in the wing, and the level of expression per cell was also lower in the

The author is in the Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA. E-mail: dstern@princeton.edu

haltere. That is, less Dpp is produced in the haltere. Remarkably, they also found that the receptor Tkv is expressed in different patterns and amounts in the wing and in the haltere. In the central part of the wing, Tkv expression is low, allowing the Dpp protein to move far from its source and creating two peaks of Dpp signaling on either side of the Dpp source. In the haltere, by contrast, all cells express high levels of Tkv, thus trapping Dpp close to the source and creating a narrow band of cells that respond to the Dpp signal. The result of all of this is that, relative to the haltere, more cells in the wing are exposed to the Dpp signal and they proliferate more than haltere cells.

Evolution appears to have hijacked an existing mechanism of growth control when flies evolved halteres. *Ubx* directs halteres to be smaller than wings by regulating multiple points in the Dpp pathway. It is remarkable that both the amount of the growth signal and the distance it is allowed to travel have come under the control of *Ubx*. Crickmore and Mann note that similar changes would provide an elegant mechanism for altering organ shape and size in different species. There is already evidence that changes in the expression of *Bmp4*, a relative of the *dpp* gene, in Darwin's finches are correlated with changes in the shape of the finches' beaks (*6*).

So how general is this mechanism? Molecules such as Dpp that transmit information through a field of cells in a graded manner are called morphogens. It is easy to imagine, given the data presented by Crickmore and Mann, that evolutionary alterations in the production and transport of morphogens through fields of cells could explain much of the geometric diversity observed by D'Arcy Thompson. Whether this is in fact the usual manner in which organ shape and overall shape evolve remains to be seen.

One particular difficulty is how this phenomenon scales to larger sizes. Morphogens tend to act over distances of tens of cells. It is difficult to imagine that tweaking a morphogen signal can lead to the differences between a fruit fly wing and a butterfly wing, or between a mouse and an elephant. Perhaps larger animals make larger organs by tinkering with morphogen gradients, or perhaps they generate new domains of morphogen activity. They may also have adopted an entirely different process that communicates with mechanisms controlling overall body size. There has recently been considerable progress in understanding how the insulin signaling pathway and other hormones (*7*, *8*) control body size in animals, but there is as yet little clarity about how morphogens and hormones intersect. It is nonetheless clear from the work of Crickmore and Mann that modification of the production and transport of morphogens may provide evolution with at least one powerful and flexible tool for altering organism shape.

References

1. D. W. Thompson, *On Growth and Form* (Cambridge Univ. Press, Cambridge, 1917).
2. J. S. Huxley, *Problems of Relative Growth* (Methuen, London, 1932).
3. M. A. Crickmore, R. S. Mann, *Science* **313**, 63 (2006); published online 2 June 2006 (10.1126/science.1128650).
4. E. B. Lewis, *Am. Zool.* **3**, 33 (1963).
5. S. D. Weatherbee, G. Halder, J. Kim, A. Hudson, S. Carroll, *Genes Dev.* **12**, 1474 (1998).
6. A. Abzhanov, M. Protas, B. R. Grant, P. R. Grant, C. J. Tabin, *Science* **305**, 1462 (2004).
7. H. F. Nijhout, *Dev. Biol.* **261**, 1 (2003).
8. J. W. Truman *et al.*, *Science* **312**, 1385 (2006).

10.1126/science.1130785

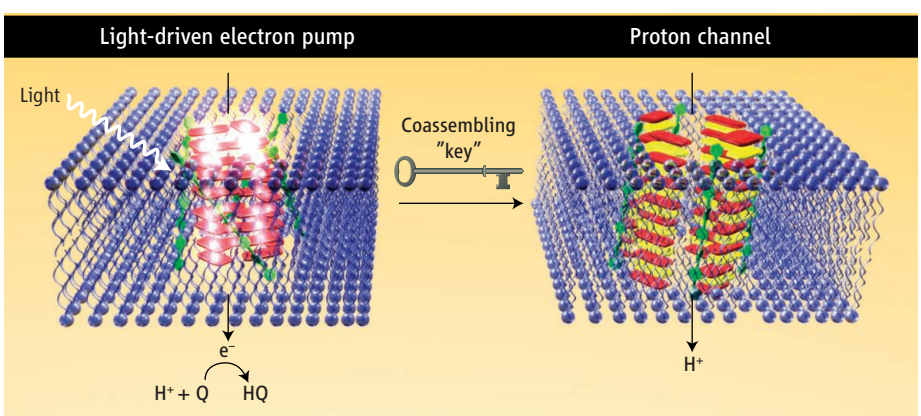
CHEMISTRY

From Electron Pump to Proton Channel

Artificial chemical systems can mimic biological ones and pump electrons across membranes in response to light. They can also be engineered to exhibit other desirable features.

Kazushi Kinbara and Takuzo Aida

Hydroelectric power generation uses the potential energy stored in dammed water. When an upstream gate is opened, the downstream flow of water drives a turbine to generate electric power. Such energy storage and release systems are also important at nanometer dimensions in cells and biological machines, where molecules control all the events involved. On page 84 of this issue, Bhosale *et al*. extend these concepts to a fully synthetic system (*1*). By integrating a self-assembled molecular gate into a vesicular membrane, they demonstrate an elegant proton-consuming and -refilling system (*1*). When the gate is closed, it serves as a light-driven electron transport pathway; when the gate is open, it forms an ion channel. Addition of a molecule serves as a key for opening the gate (see the figure).



Generation and release of a concentration gradient of protons. Transformation of a light-driven electron pump (**left**) into an ion channel (**right**) is triggered by an externally added molecule that serves as a key. Q, quinone derivative; HQ, hydroquinone derivative.

CREDIT: KAZUSHI KINBARA

At a molecular level, energy can be gained from fluxes of substances along their concentration gradient. Many biological events, including flagellum rotation, signal transduction, and muscle contraction, make use of this mechanism to generate their driving energy. In biological systems, protons, Na^+, K^+, or Ca^{2+} ions are concentrated on one side of a biomembrane by the action of chemical "pumps." When certain channels are constructed in the biological membrane, unidirectional flows of these ions take place. Some biological membranes, such as thylakoid membranes, contain light-harvesting chromophore arrays that enable the use of light energy for generating a concentration gradient of protons.

Ever since Kunitake (*2*) and Ringsdorf (*3*) pioneered the development of artificial liposomes (vesicular assemblies of amphiphilic molecules), researchers have tried to realize energy

The authors are in the Department of Chemistry and Biotechnology, School of Engineering, and the Center for NanoBio Integration, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. E-mail: aida@macro.t.u-tokyo.ac.jp

storage and use in biologically inspired artificial systems. One of the essential steps toward this goal is the construction of ion channels in artificial membranes. To control the local concentrations of ions in cells or organisms, biological systems form various ion channels in response to external factors or environmental changes such as ligand binding. Tabushi *et al.* (*4*) used β-cyclodextrin as a scaffold. Since then, a number of artificial ion channels have been developed, including channels made from peptide-core dendrimers (*5*). When channels are decorated with photochromic molecules, fluxes of ions can be changed in response to light stimuli (*6*).

To realize an artificial energy storage and retrieval system, the artificial membranes must be equipped with an ion-pumping mechanism. Inspired by biological light-harvesting membranes, some early attempts used chlorophyll for the construction of a light-driven electron pump in artificial membranes. Later, Calvin and co-workers created a fully artificial system by incorporating a ruthenium bipyridine complex into an artificial membrane as a sensitizer (*7*).

More recently, Moore and co-workers have created a sophisticated chemical pumping system through the use of a photoactive dyad with covalently linked electron donor and acceptor units (*8*). Upon photoirradiation, the dyad, which is embedded in a vesicular membrane, forms a charge-separated state. The electron is produced at one end of the dyad, is transferred to another electron acceptor molecule in the membrane, and finally generates a proton through reduction of quinone derivatives, followed by reoxidation. Thus, the system can use light energy to generate and locally accumulate protons. Because the photoactive membrane is devoid of any mechanism for the construction of channels, Moore and co-workers have integrated ATP synthase, which uses generated protons to convert adenosine diphosphate (ADP) into energy-storing adenosine triphosphate (ATP), into the membrane (*9*).

In contrast with these earlier examples, Bhosale *et al.* have realized a one-way switch from the generation to the release of a concentration gradient of protons in response to an externally added molecule, which acts as a key. The authors take advantage of the dynamic character of supramolecular systems, which are held together through comparatively weak intermolecular interactions. In their vesicular membrane, an electron-deficient naphthalene diimide (NDI) derivative self-assembles (with the assistance of four molecules of a rodlike oligophenylene with peptide branches) to form π-stacked helical arrays. Upon photoexcitation with visible light (638 nm), the π-stacked arrays accept an electron from a sacrificial electron donor placed outside of the vesicular spheres. The array then rapidly transports the acquired electron into the vesicles, which contain an electron-accepting quinone derivative. As a result of this transmembrane electron transport, the quinone is reduced to form a hydroquinone. Repetition of this photochemical process leads to the consumption of protons in the vesicles.

To convert this pump into an ion channel, Bhosale *et al.* took advantage of the ability of supramolecular assemblies to reorganize. By adding an electron-rich dialkoxynaphthalene derivative to the system, they were able to transform the NDI arrays into a channel-like hollow coassembly (*10*). A transmembrane proton (hydroxide ion) flux through the resulting channels results. At the same time, the electron transport function embedded in the membrane becomes disabled, because the coassembly is photochemically inactive.

Thus, Bhosale *et al*. have succeeded in constructing a light-driven electron pump that can easily be converted into a chemically driven proton channel. However, it cannot yet be switched back into pump mode. Future studies of this highly interesting system may include attempts to integrate a reversible switch between the pump and the channel modes of the system.

References

1. S. Bhosale *et al.*, *Science* **313**, 84 (2006).
2. T. Kunitake, Y. Okahata, *J. Am. Chem. Soc.* **99**, 3860 (1977).
3. H.-H. Hub, B. Hupfer, H. Koch, H. Ringsdorf, *Angew. Chem. Int. Ed.* **19**, 938 (1980).
4. I. Tabushi, Y. Kuroda, K. Yokota, *Tetrahedron Lett.* **23**, 4601 (1982).
5. V. Percec *et al*., *Nature* **430**, 764 (2004).
6. L. Lien, D. C. J. Jaikaren, Z. Zhang, G. W. Wooley, *J. Am. Chem. Soc.* **118**, 12222 (1996).
7. W. Ford, J. W. Otvos, M. Calvin, *Nature* **274**, 507 (1978).
8. G. Steinberg-Yfrach *et al*., *Nature* **385**, 239 (1997).
9. G. Steinberg-Yfrach *et al*., *Nature* **392**, 479 (1998).
10. P. Talukdar et al., *J. Am. Chem. Soc.* **127**, 6528 (2005).

10.1126/science.1130124

CELL BIOLOGY

Stressed Cells Cope with Protein Overload

Stress on eukaryotic cells causes build-up of unfolded proteins in the endoplasmic reticulum. To protect itself, the cell degrades mRNA for ER proteins.

David Ron

Most secreted and membrane proteins of eukaryotic cells undergo maturation steps in the lumen of the endoplasmic reticulum (ER). Upon translation from an RNA template, nascent polypeptides are translocated into the ER lumen where they are chemically modified and folded into mature proteins. Consequently, demands made on the ER's protein folding machinery must be monitored to avoid undue ER stress. The cell's unfolded protein response (UPR) mechanisms attempt to balance the demand with the capabilities of the protein folding machinery. Indeed, failure to execute this response is associated with cellular dysfunction and disease (*1*). On page 104 of this issue, Hollien and Weissman describe a new mechanism by which cells experiencing ER stress selectively degrade messenger RNAs (mRNAs) that would otherwise encode polypeptides destined for the ER lumen (*2*). This hitherto unanticipated aspect of the UPR highlights the importance of posttranscriptional remodeling in adapting to ER stress.

Until now, the UPR was believed to have two major effector mechanisms: reducing translation initiation and increasing the transcription of genes that contribute broadly to ER function (*3*, *4*). The translational arm of the UPR constitutes a short-term adaptation that reduces the load of newly synthesized polypeptides entering the ER lumen. The transcriptional arm constitutes a long-term adaptation that increases the capacity of the organelle to handle unfolded proteins. Hollien and Weissman now add a third effector—the degradation of mRNAs encoding secreted and membrane proteins. The rapid and selective elimination of such mRNAs is likely to complement translational repression and further reduce the unfolded protein load on the stressed ER. However, its consequences may extend to the later phases of the response, because degradation of a subset of (membrane-associated) mRNAs will free up ER-associated translational and translocational machinery to accommodate newly synthesized mRNAs induced by the UPR's transcriptional arm. Translational repression plays a similar role, because it selectively targets the initiation of protein synthesis. This increases the pool of free ribosomal subunits that

The author is at the Skirball Institute, New York University Medical Center, 540 First Avenue, New York, NY 10016, USA. E-mail: ron@saturn.med.nyu.edu

are available to translate the newly made mRNAs, later, upon reversal of the translational repression (*4*). According to this theory, mRNA degradation and translational repression synergistically contribute to reprogramming the ER in stressed cells (see the figure).

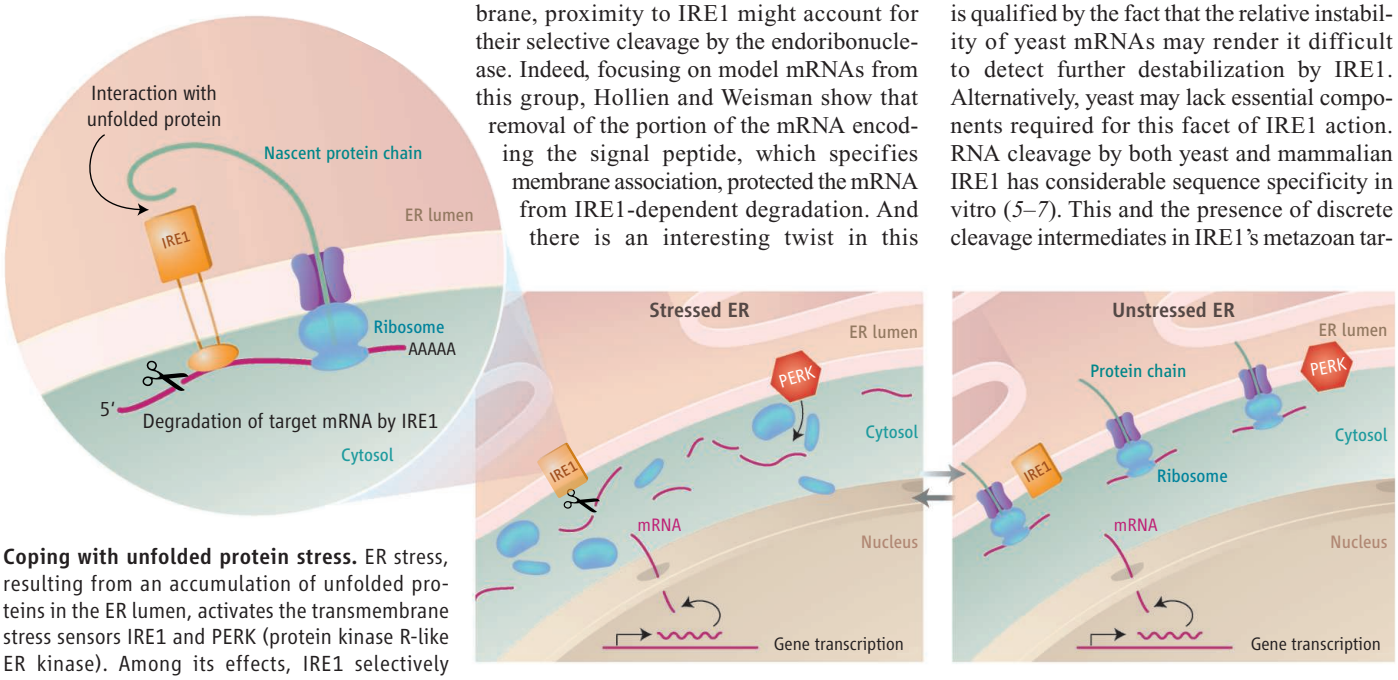
Remarkably, the regulated degradation of mRNAs in ER-stressed cells was found to depend on inositol-requiring enzyme–1 (IRE1), the phylogenetically oldest component of the UPR. IRE1 is a bimodular protein that traverses the ER membrane. IRE1's lumenal, stress-sensing, portion responds to an imbalance between unfolded proteins and the ER chaperones that catalyze protein folding by activating its cytoplasmic, effector portion. The latter entails transautophosphorylation in IRE1's cytoplasmic kinase domain and the linked unmasking of an endoribonucleolytic activity. In yeast, activated IRE1 cleaves preformed, inactive, *HAC1* (*homologous to ATF/CREB 1*) mRNA at two precise points. Religation of the 5′ and 3′ fragments leads to a spliced *HAC1* mRNA that encodes a potent activator of genes that function in the UPR (*5*).

In metazoans, too, IRE1 activates gene expression by precise splicing of *XBP-1* (*X-box binding protein 1*) mRNA (*6*, *7*), but, as Hollien and Weissman show, IRE1 also has an unexpected, destructive effect on other mRNAs. Comprehensive analysis of mRNAs expressed in cultured *Drosophila melanogaster* S2 cells revealed that depletion of IRE1 or XBP-1 (by RNA interference) inhibited the expression of a set of genes normally activated by ER stress. This was expected, given XBP-1's role in mediating the transcriptional effects of IRE1 activation. Surprisingly, analysis of genes whose expression normally declines during ER stress revealed a subset that depends on IRE1 but not on XBP-1 for their decline. Furthermore, decreased stability, rather than reduced transcription, accounted for this IRE1-dependent change in mRNA level.

This pool of mRNAs was enriched in those encoding secreted proteins. Because these are synthesized in association with the ER membrane, proximity to IRE1 might account for their selective cleavage by the endoribonuclease. Indeed, focusing on model mRNAs from this group, Hollien and Weissman show that removal of the portion of the mRNA encoding the signal peptide, which specifies membrane association, protected the mRNA from IRE1-dependent degradation. And there is an interesting twist in this story—neither proximity to the ER, nor the nucleic acid sequence of the mRNA, was sufficient to specify degradation. The signal peptide, which directs the translating ribosome to the ER membrane, also promotes translocation of the nascent polypeptide into the ER lumen. Altering the frame in which the mRNA is read, and thereby the sequence of the translocated polypeptide, abolished IRE1-mediated degradation.

These findings implicate signaling across the ER membrane in mRNA degradation, because the nascent polypeptide is in the lumen and the mRNA is in the cytosol. Recent insight from the structure of IRE1 suggests how this might come about, because its lumenal domain appears to contain a deep groove suited to interact directly with unfolded proteins (*8*). Therefore, in one particularly interesting scenario, engagement of IRE1 with the nascent polypeptide on the lumenal side could specify the destruction of its encoding mRNA in the cytosol of stressed cells (see the figure).

IRE1 is the oldest component of the UPR, and the nonconventional splicing of its effector mRNAs HAC1 or XBP-1 is conserved throughout eukaryotes. Nonetheless, IRE1 function has evolved to include recruitment of metazoan-specific components that connect the UPR with other important signaling pathways (*9*, *10*). The IRE1-dependent degradation of mRNAs also appears to be a metazoan embellishment because there is no evidence for its existence in yeast (*11*). This conclusion is qualified by the fact that the relative instability of yeast mRNAs may render it difficult to detect further destabilization by IRE1. Alternatively, yeast may lack essential components required for this facet of IRE1 action. RNA cleavage by both yeast and mammalian IRE1 has considerable sequence specificity in vitro (*5*–*7*). This and the presence of discrete cleavage intermediates in IRE1's metazoan targets mRNAs (*2*) suggest the coevolution of hotspots for IRE1 action on the mRNAs. Either way, the apparent differences between unicellular organisms and metazoans in regard to this new aspect of the UPR emphasize the need for more work on the mechanistic aspects of this example of regulated mRNA instability.



Coping with unfolded protein stress. ER stress, resulting from an accumulation of unfolded proteins in the ER lumen, activates the transmembrane stress sensors IRE1 and PERK (protein kinase R-like ER kinase). Among its effects, IRE1 selectively cleaves mRNAs encoding secreted proteins. PERK transiently inhibits the initiation of protein synthesis and liberates free ribosomal subunits. These actions reprogram the ER-associated translational apparatus to accommodate changes in gene transcription in stressed cells. **(Inset)** A hypothetical model showing that in stressed cells, an unchaperoned and unfolded nascent polypeptide that has been translocated across the ER membrane engages the lumenal domain of IRE1 and activates it to cleave the encoding mRNA on the cytosolic side of the ER membrane. The cleaved mRNA is then further degraded by other ribonucleases.

References

1. R. J. Kaufman, *J. Clin. Invest.* **110**, 1389 (2002).
2. J. Hollien, J. S. Weissman, *Science* **313**, 104 (2006).
3. C. Patil, P. Walter, *Curr. Opin. Cell Biol.* **13**, 349 (2001).
4. H. P. Harding, M. Calfon, F. Urano, I. Novoa, D. Ron, *Annu. Rev. Cell Dev. Biol.* **18**, 575 (2002).
5. C. Sidrauski, P. Walter, *Cell* **90**, 1031 (1997).
6. H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, *Cell* **107**, 881 (2001).
7. M. Calfon *et al.*, *Nature* **415**, 92 (2002).
8. J. J. Credle, J. S. Finer-Moore, F. R. Papa, R. M. Stroud, P. Walter, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18773 (2005).
9. F. Urano *et al.*, *Science* **287**, 664 (2000).
10. C. Hetz *et al.*, *Science* **312**, 572 (2006).
11. K. J. Travers *et al.*, *Cell* **101**, 249 (2000).

10.1126/science.1130469

PLANT SCIENCE

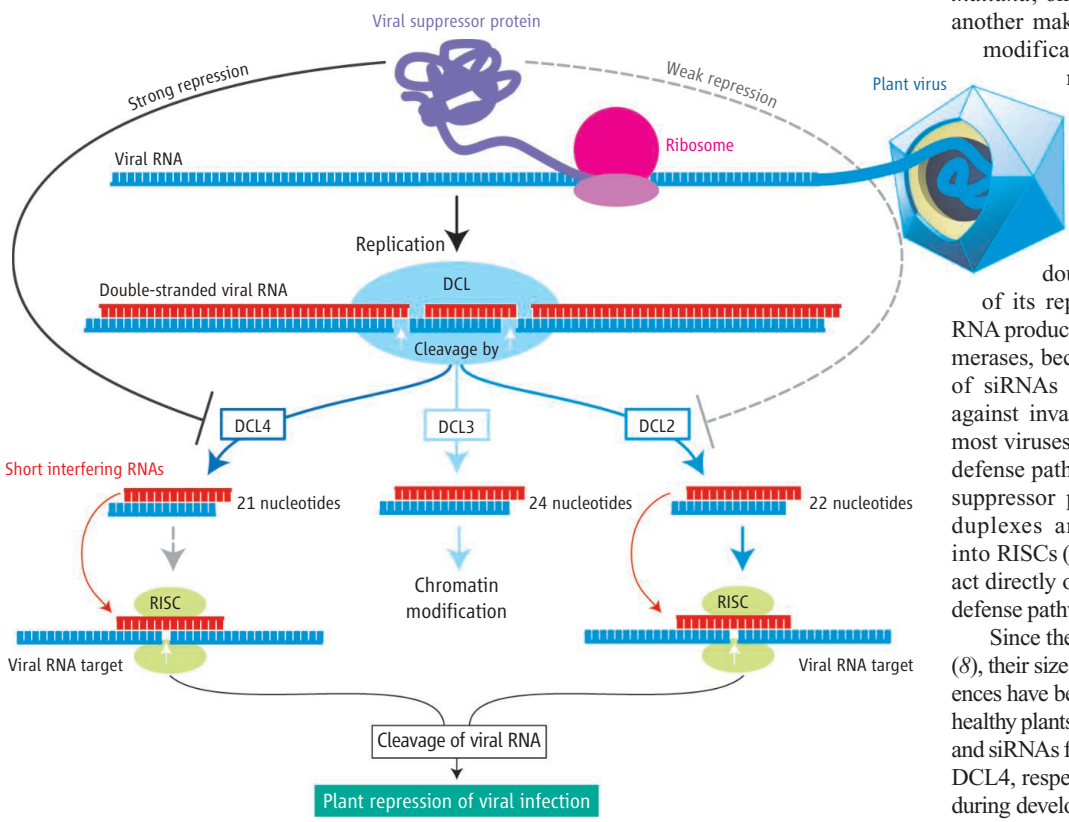
Viruses Face a Double Defense by Plant Small RNAs

Plants fight viral infections with enzymes that digest viral RNA, but viruses retaliate with proteins that suppress these enzymes. To boost their antiviral response plants deploy enzymes with redundant functions.

Peter M. Waterhouse and Adriana F. Fusaro

All organisms, from bacteria to mammals, can be infected by viruses. Bacteria counter infection with endonucleases that cleave viral DNA. Mammals fight infection with antibodies and lymphocytes that are adapted to specific viral antigens. They also employ nonadaptive defenses, such as producing interferons that block viral replication and stimulate the host immune response, and launching an editing attack on viral RNA (adenosine deaminases acting on RNA). However, none of these defense mechanisms has been found in plants. Instead, plants appear to rely heavily on an adaptive RNA degradation system (most plant viruses have RNA genomes) similar to that mediating RNA interference (a gene silencing mechanism) in animals. On page 68 of this issue, Deleris *et al.* (*1*) reveal new molecular insights into how viruses both trigger and suppress this antiviral RNA silencing mechanism in host plants.

The plant antiviral response involves Dicers, enzymes similar to ribonuclease III, which cleave viral double-stranded RNA into smaller pieces (~21 to 24 nucleotides in length) called short interfering RNA (siRNA) duplexes (see the figure). One strand of each duplex is incorporated into a large ribonucleoprotein complex called the RNA-induced silencing complex (RISC). Guided by the associated siRNA, RISC recognizes and destroys complementary target viral RNAs (*2*). In animals, and probably all multicellular eukaryotes, Dicer is also involved in the production of microRNAs, which are very similar to siRNAs but generated from endogenous, partially self-complementary precursor RNAs. These microRNAs direct RISCs to silence the expression of key genes during important developmental transitions (*3*). From the available complete genome sequences, it seems that whereas metazoans and fungi have only one or two Dicer genes, plants have multiplied and modified them over the past billion years to produce a basic set of at least four different Dicer-like (DCL) variants of the enzyme (*4*). In the model plant *Arabidopsis thaliana*, one DCL produces microRNAs and another makes siRNAs that direct chromatin modification. Deleris *et al.* show that the remaining two DCLs combine to provide an antiviral defense system that itself becomes a target of viral counterdefense mechanisms.



Plant antiviral defense. Pathways directed by Dicer-like (DCL) enzymes interact with viral RNAs and are affected by viral suppressor proteins. Viruses are repressed mainly through DCL4. However, Turnip crinkle virus encodes a suppressor protein that inhibits DCL4 activity. In this situation, the 22-nucleotide siRNAs produced by DCL2 become the major directors of antiviral activity.

When a virus infects a plant, the partially self-complementary structures of its genomic RNA, double-stranded RNA intermediates of its replication, and/or double-stranded RNA produced from the viral RNA by host polymerases, become substrates for the production of siRNAs (*5*). These siRNAs guide RISCs against invading viral RNA. Unsurprisingly, most viruses have evolved ways to combat this defense pathway. Many viral genomes encode suppressor proteins that bind to the siRNA duplexes and prevent their incorporation into RISCs (*6*). Others produce suppressors that act directly on the enzymes or cofactors of this defense pathway (*7*).

Since the first detection of siRNAs in plants (*8*), their sizes and the relevance of the size differences have been debated. The current view is that healthy plants generate 21-nucleotide microRNAs and siRNAs from their own RNA, with DCL1 and DCL4, respectively, to regulate gene expression during development. DCL3 generates 24-nucleotide siRNAs to modify chromatin. But the size profile of viral siRNAs varies depending on the virus. Many viruses induce the production of mostly 21-nucleotide siRNAs. Others cause the plant to make a mix of 21- and 24-nucleotide siRNAs. And other viruses trigger the generation of predominantly 24-nucleotide or exclusively 22-nucleotide viral siRNA species. So, why are the different viruses treated differently by plants? Deleris *et al.* show that the answer has a lot to do with DCL redundancy and the action of viral suppressor proteins on DCL-directed activity.

Arabidopsis infected with *Tobacco rattle virus* (TRV) generates virus-specific 21- and 24-nucleotide siRNAs, and when infected by *Turnip crinkle virus* (TCV), it produces only 22-nucleotide siRNAs. However, these size profiles change in different *Arabidopsis Dcl*-mutant

The authors are at CSIRO Plant Industry, Canberra, ACT 2601, Australia. E-mail: peter.waterhouse@csiro.au

Published by AAAS

backgrounds. The TCV 22-nucleotide siRNAs produced by DCL2 direct destruction of the virus, whereas the TRV 21- and 24-nucleotide species are generated by DCL4 and DCL3, respectively, with only the 21-nucleotide siRNAs directing destruction of the virus. Why is this and how is it done? Deleris *et al*. show that TCV RNA would be targeted by DCL4-produced siRNAs if it were not for the virus expressing a suppressor protein that blocks DCL4 activity. The contrary situation exists for TRV. If DCL4 activity is eliminated from the plant by mutation, RISC is guided against TRV by 22-nucleotide siRNAs produced by DCL2. Both viruses replicate to greater levels in a double *dcl2-dcl4* mutant, showing that even though the siRNAs from one of the DCLs predominates, both DCLs are producing siRNAs that restrict the virus. However, DCL4 appears to be the first line of antiviral defense, with DCL2-mediated activity coming to the fore when DCL4 is deactivated by a viral suppressor protein.

Have plants duplicated DCLs as a response to viral suppressors? Mammals have only one Dicer plus a suite of other viral defense systems, and it is currently debated whether their RNA interference pathway has an antiviral role (*9*). Insects have fewer antiviral defense systems than mammals but have two Dicers, and a number of insect viruses encode suppressors that inhibit these Dicers or their pathways (*10*, *11*). Plants such as rice and poplar have even more DCLs than *Arabidopsis*, which are probably involved in antiviral defense (*4*). Other plant viruses, such as citrus tristeza, encode three different suppressor proteins to shut down the DCL-mediated defense responses (*12*). All this suggests that the RNA silencing-based antiviral defense pathway is increasingly important to eukaryotes that have fewer alternative antiviral protection systems, and that there is an escalating arms race between viruses and their plant hosts.

The study raises the question: How can a viral suppressor protein inhibit the activity of a specific DCL in the production of viral siRNAs? In this case, the answer may reside in the RNA-binding capacity of the viral suppressor protein P38, which can bind to long as well as short (21- to 26-nucleotides) double-stranded RNA (*13*). P38 may be competing with DCL4, blocking DCL4's access to long double-stranded viral RNA molecules. Alternatively, P38 might directly interact with DCL4, affecting its activity, but not that of DCL2.

The complexity of small RNA–mediated pathways in plants, and the roles within them played by each of the DCLs, is becoming ever more apparent. A major challenge will be to determine the subcellular localization of the pathway components and details of their trafficking to these locations. DCL1, DCL3, and DCL4 have all been detected exclusively in the nucleus (*14*, *15*), which is consistent with the production of microRNAs and siRNAs from host transcripts. However, the demonstration by Deleris *et al.* that all of these enzymes have access to RNAs from viruses thought not to enter the nucleus suggests that either DCLs or viral RNA substrates are being trafficked between the cytoplasm and nucleus.

References

1. A. Deleris *et al*., *Science* **313**, 68 (2006); published online 1 June 2006 (10.1126/science.1128214).
2. Y. Tomari, P. D. Zamore, *Genes Dev.* **19**, 517 (2005).
3. D. P. Bartel, *Cell* **116**, 281 (2004).
4. R. Margis *et al.*, *FEBS Lett.* **580**, 2442 (2006).
5. O. Voinnet, *Nat. Rev. Genet*. **6**, 206 (2005).
6. L. Lakatos *et al.*, *EMBO J.* 10.1038/sj.emboj.7601164 (25 May 2006).
7. M.Pazhouhandeh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1994 (2006).
8. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (1999).
9. B. R. Cullen, *Nat. Immunol.* **7**, 563 (2006).
10. H. Li, W. X. Li, S. W. Ding, *Science* **296**, 1319 (2002).
11. X-H. Wang *et al.*, *Science* **312**, 452 (2006).
12. R. Lu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15742 (2004).
13. Z. Merai *et al.*, *J. Virol.* **80**, 5747 (2006).
14. Z. Xie *et al.*, *PLoS Biol.* **2**, E104 (2004).
15. A. Hiraguri *et al.*, *Plant Mol. Biol.* **57**, 173 (2005).

10.1126/science.1130818

CHEMISTRY

Self-Assembly in Action

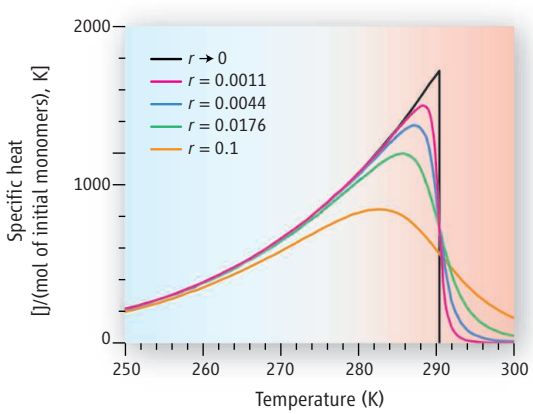
Virgil Percec, Goran Ungar, Mihai Peterca

Growth of supramolecular structures is in some respects analogous to a phase transition and thus can be described by familiar thermodynamic models.

In 1983, Aaron Klug showed that tobacco mosaic virus and other biological complexes are self-assembling, self-nucleating, and self-checking (*1*). Chemists soon found that the same principles apply to the self-assembly of complex synthetic systems (*2*). Today, synthetic complexes can mimic helical rod-like and spherical biological systems (*3*–*5*). However, although their structure and functions have received much attention, the mechanism of their formation remains poorly understood (*6*). On page 80 of this issue, Jonkheijm *et al.* (*7*) report on a detailed study of the process of molecular self-assembly in solution and show it to be similar to a phase transition commonly associated with ordered solids or liquid crystals.

In the past, there have been rare but notable studies of the different stages of the hierarchical process of supramolecular association. For example, Lehn and co-workers used transmission electron microscopy to study complexes of chiral (left- and right-handed) tartaric acid derivatives (*8*). These complexes first form small self-nucleated clusters; the clusters then grow into helical filaments that in turn assemble into helical fibers. However, transmission electron microscopy can only resolve initial clusters above a certain size.

Jonkheijm *et al.* have overcome this problem by using optical analytical techniques combined with an appropriate theoretical thermodynamic model of activated aggregation. They studied a paraffin solution of a chiral rigid-rod chromophore with a hydrogen-bonding group at one end and a bulky flexible group, known to promote columnar self-assembly, at the other. Below a certain temperature, the molecules aggregate to form supramolecular helical fila-



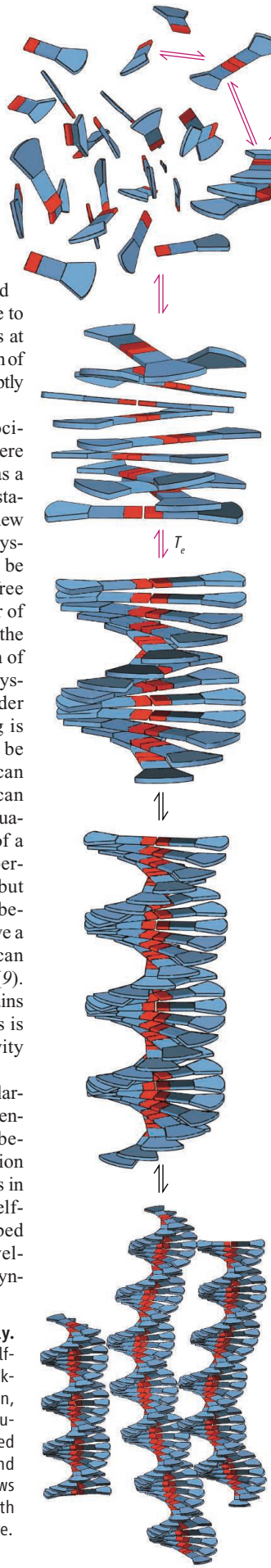
How to spot a phase transition. Behavior of the specific heat of living polymerization as a function of temperature for different initiator concentrations *r*. (The specific heat is proportional to the number of broken bonds per unit increase in temperature.) As *r* tends toward 0, the polymerization behaves as a second-order phase transition, and the specific heat shows a discontinuity at the critical temperature. The curve becomes rounded with increasing *r*. The theoretical curves shown here [adapted from (*10*)] may be compared with the experimental specific heat peak in Fig. 2J of the report by Jonkheijm *et al.*

V. Percec and M. Peterca are in the Roy & Diana Vagelos Laboratories, Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104, USA. E-mail: percec@sas.upenn.edu peterca@sas.upenn.edu G. Ungar is in the Department of Engineering Materials, University of Sheffield, Sheffield S1 3JD, UK. E-mail: g.ungar@shef.ac.uk

ments (see the figure at right). Using a range of optical techniques, the authors were able to detect and distinguish different states of molecular aggregation as a function of temperature under equilibrium conditions. They identified a hierarchy of aggregated structures, from individual hydrogen-bonded dimers at the highest temperature to clusters of helical nanofilaments at the lowest temperature. The fraction of molecules in aggregates rose abruptly below a critical temperature T_e.

Nucleation is usually associated with phase transitions, where the transformation is hindered as a result of the thermodynamic instability of a small domain of the new phase (the nucleus). Thus, in crystallization, a small crystal must be formed first, with its overall free energy higher than that of either of the two bulk phases because of the comparatively high contribution of surface energy. In the case of crystallization and other first-order transitions, some supercooling is needed before the barrier can be overcome and the transition can proceed. Nucleation barriers can cause highly nonequilibrium situations, such as the glassy state of a supercooled liquid. Highly supercooled liquids are common, but superheated crystals are rare, because crystals almost always have a free surface on which melt can form with virtually no barrier (*9*). The fact that large ordered domains are more stable than small ones is the general cause of cooperativity in phase transitions.

Despite the apparent dissimilarity of the two situations, fundamental analogies can be drawn between supramolecular aggregation in solutions and phase transitions in condensed matter. Solution self-assembly is commonly described by thermodynamic models developed for the polymerization of synthetic or biological macromolecules (*10*). In equilibrium freely associating systems, also known as isodesmic systems, there is an equilibrium cluster size distribution for any given temperature. Such systems display a rounded sigmoidal curve if the fraction of aggregated molecules is plotted against temperature. This is the case, for example, in the micellar aggregation of surfactants in solution or in some cases of living polymerization without a reaction terminator.



Nucleated helical self-assembly. Schematic representation of the self-assembly process described by Jonkheijm *et al.*, involving dimerization, nonhelical aggregation, helical nucleation, and helical growth. Red arrows indicate assembly up to and including nucleation; black arrows refer to the subsequent growth process. T_e is the critical temperature.

However, if the aggregation involves an activated, high–free energy state, such as in thermally or chemically initiated polymerization, then the sigmoidal curve changes into a curve more resembling a phase transition. The lower the equilibrium constant of the activated state (that is, the less abundant it is), the higher the cooperativity and the sharper the transition (see the first figure).

In the case of polymerization reactions, the temperature equivalent to the polymerization transition is known as the ceiling temperature; above it, the polymer depolymerizes to monomer while below it, the monomer polymerizes (*11*). This situation is in some way analogous to the boiling point of a liquid. Just as water will evaporate from an uncovered bowl well below the boiling point, so can a polymer depolymerize well below its ceiling temperature. (A few house owners have discovered this to their chagrin when they found that their polystyrene loft insulation had disappeared over time.)

What is the activated state in the self-assembly process described by Jonkheijm *et al.*, and why is the transition sharp and highly cooperative? Thanks to the application of different optical techniques, each of which detects a different aspect of the self-assembly process, the authors identified a small nonhelical aggregate of hydrogen-bonded dimers as the high-energy nucleus (see the second figure). Once helical order is established, the energy drops and, provided the temperature is below T_e, the supramolecular filament grows. The equilibrium constant of the activated species is low (10^{-4}), as determined from the fit of the circular dichroism data to the thermodynamic model. Thus, the process is highly cooperative.

Some self-assembly processes are activated and cooperative and some are not; there are also those where the first association step is more favored than the subsequent ones. In the latter case, the process is anticooperative. For example, Arnaud *et al*. found that for the formation of columnar filaments of a perylene derivative in water, the equilibrium constant for dimerization is 8 times that for the subsequent polymerization steps (*12*). Such a situation generally tends to impair the stability of the supramolecular aggregate.

The above discussion relates to self-nucleation (commonly known as homogeneous nucleation in the field of crystallization). There is a whole alternative world out there of directed self-assembly nucleated on designed surfaces (*13*) (heterogeneous nucleation), which includes such diverse areas as biomineralization and nanofabrication of electronic and photonic components.

The work of Jonkheijm *et al*. gives us details of the early and advanced stages of a highly cooperative process of self-assembly involving homogeneous nucleation. Additional studies along these lines will help the self-assembly of complex functional nanosystems to approach the level of structural and functional precision encountered in biological systems (*14*, *15*).

References

1. A. Klug, *Angew. Chem. Int. Ed.* **22**, 565 (1983).
2. J.-M. Lehn, *Angew. Chem. Int. Ed.* **27**, 89 (1988).
3. V. Percec *et al.*, *Nature* **430**, 764 (2004).
4. G. Ungar, Y. Liu, X. Zeng, V. Percec, W.-D. Cho, *Science* **299**, 1208 (2003).
5. H. Engelkamp, S. Middelbeek, R. J. M. Nolte, *Science* **284**, 785 (1999).
6. K. van Workum, J. F. Douglas, *Macromol. Symp.* **227**, 1 (2005).
7. P. Jonkheijm, P. van der Schoot, A. P. H. J. Schenning, E. W. Meijer, *Science* **313**, 80 (2006).
8. T. Gulik-Krzywicki, C. Fouquei, J.-M. Lehn, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 163 (1993).
9. R. W. Cahn, *Nature* **323**, 668 (1986).
10. J. Dudowicz, K. F. Freed, J. F. Douglas, *J. Chem. Phys.* **113**, 434 (2000).
11. K. J. Ivin, *J. Polym. Sci.* **38**, 2137 (2000).
12. A. Arnaud *et al.*, *Angew. Chem. Int. Ed.* **43**, 1718 (2004).
13. J. Aizenberg, A. J. Black, G. M. Whitesides, *Nature* **398**, 495 (1999).
14. H. Bayley, P. S. Cremer, *Nature* **413**, 226 (2001).
15. T. Douglas, M. Young, *Science* **312**, 873 (2006).

10.1126/science.1129512

RETROSPECTIVE

R. Bruce Merrifield (1921–2006)

Richard A. Lerner

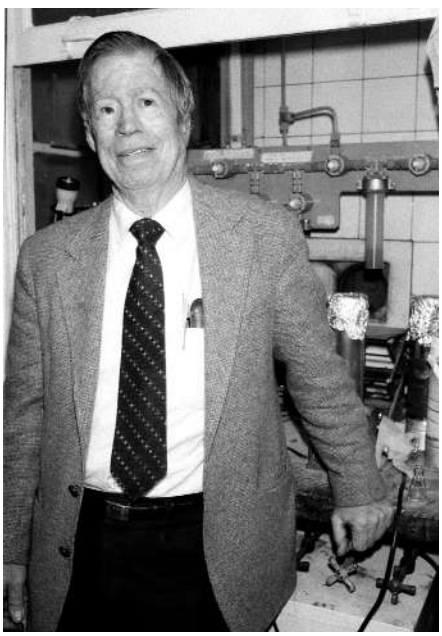
R. Bruce Merrifield, who developed the method of solid-state synthesis, had a talent for simplicity and practicality and was stoic in the face of adversity.

On 14 May 2006, Robert Bruce Merrifield died at the age of 84 after a long illness. He was an unusual chemist who taught the chemistry community a new way to think about organic synthesis.

Merrifield grew up in California and after a rather itinerant educational experience, he received his Ph.D. in chemistry from the University of California at Los Angeles in 1949. He then moved on to the Rockefeller Institute for Medical Research (later Rockefeller University), where he remained for the rest of his life. In 1963, he published a seminal paper that launched the method of solid-state synthesis, in which reactions are carried out on a solid support rather than in solution. The centerpiece of his insight was that important biological molecules such as peptides and proteins were linear polymers and that one could thus attach one end to a matrix during their synthesis. The method was initially used for the synthesis of peptides and proteins; more recently, it has found wide use in the synthesis of oligonucleotides, carbohydrates, and even for the construction of complicated organic molecules where the individual steps in a reaction series are less repetitive and thus chemically more diverse.

Initially, the method of solid-state synthesis received less-than-enthusiastic support from chemists who were used to a rigorous analysis of the products from each step of a reaction series. Indeed, at that time and today, the body of organic chemistry grows by building a knowledge base of chemical reactivity. This base owes its vigor to the annotation of thousands of thoroughly analyzed attempts at chemical transformations. Both successes and failures are dutifully recorded. By contrast, solid-state methods did not allow analysis at each step; only the finished product could be analyzed after it was cleaved from the support.

But Merrifield had other fish to fry: He wanted a practical solution for the synthesis of important large biological polymers such as proteins, where one could not suffer the losses of material attendant to the need to isolate and purify the products from solution for each step in a very long reaction series. Given the funding climate we live in today, it is worth noting that in the absence of external support, his university sustained his research program and enabled his breakthrough. Even though other peptide chemists made large peptides by fragment condensation, he was alone in the belief that virtually any length of peptide or protein could be made by solid-state methods.



A talent for simplicity. Bruce Merrifield poses in his laboratory at Rockefeller University, New York, on 18 October 1984, shortly after learning that he had won the Nobel Prize for chemistry.

Arguably, the watershed event that led to the acceptance of Merrifield's process was the synthesis in 1969, together with Bernd Gutte, of the 124–amino acid–long enzyme ribonuclease A. The recovery of active enzyme with a 78% nuclease-specific activity left little doubt about the fidelity of his method.

A practical solution for peptide synthesis. In this 1969 photo, Merrifield adjusts an early version of his peptide synthesizer.

In 1984, Merrifield received the Nobel Prize in chemistry "for his development of methodology for chemical synthesis on a solid matrix." While his methods were simple, his grasp of organic reaction mechanisms was deep. In addition, he used his technology to make interesting molecules to probe biological mechanisms. Until illness kept him from the bench, he continued to study why certain peptide antibiotics had the same biological activity as their antipodes. Probably correctly, he assumed that their interaction was not diasteromeric in that their function did not depend on the interaction with chiral centers in the cell.

Despite all his accomplishments, Merrifield was an unassuming and gentle man. In later years, his method had been industrialized, and throughout the world scientists were proudly displaying the latest and greatest computerized machines for solid-state synthesis. Yet the man who had invented the method could be found in his laboratory huddled over the simple wrist shaker with which he did the original reactions. I never heard Merrifield utter an unkind remark, even when he was under assault from certain quarters in the chemical community or when he was ill. One cannot write about Merrifield without referring to his stoicism in the face of physical adversity. As a consequence of irradiation for childhood acne, Merrifield suffered a relentless series of facial tumors that only seemed to bother him because the surgery to remove them meant time away from his beloved Rockefeller University.

Merrifield is survived by his wife and coworker, Elizabeth, one son, and five daughters. Like his propensity for simplicity and practicality in the laboratory, he led an uncomplicated life. For more than 40 years, Merrifield and his friend Vincent Allfrey would together make the long commute from New Jersey to Rockefeller University in Manhattan. Bruce would drive and Vincent would read out loud from a wide variety of books. Merrifield was proud of this ritual and would tell me about what they were reading almost every time I saw him. My guess is that he relished this workday routine because it signaled to him that in an increasingly complicated world, there was still room for simple things. After all, he was widely recognized for simplifying the complicated art of organic synthesis.

10.1126/science.1130957

The author is at The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. E-mail: rlerner@scripps.edu

Global Biodiversity Conservation Priorities

T. M. Brooks,[1,2,3]* R. A. Mittermeier,[1] G. A. B. da Fonseca,[1,4] J. Gerlach,[5,6] M. Hoffmann,[1] J. F. Lamoreux,[3] C. G. Mittermeier,[1] J. D. Pilgrim,[7] A. S. L. Rodrigues[5]

The location of and threats to biodiversity are distributed unevenly, so prioritization is essential to minimize biodiversity loss. To address this need, biodiversity conservation organizations have proposed nine templates of global priorities over the past decade. Here, we review the concepts, methods, results, impacts, and challenges of these prioritizations of conservation practice within the theoretical irreplaceability/vulnerability framework of systematic conservation planning. Most of the templates prioritize highly irreplaceable regions; some are reactive (prioritizing high vulnerability), and others are proactive (prioritizing low vulnerability). We hope this synthesis improves understanding of these prioritization approaches and that it results in more efficient allocation of geographically flexible conservation funding.

Human actions are causing a biodiversity crisis, with species extinction rates up to 1000 times higher than background (*1*). Moreover, the processes driving extinction are eroding the environmental services on which humanity depends (*2*). People care most about what is close to them, so most responses to this crisis will be local or national (*3*). Thus, approximately 90% of the $6 billion of annual conservation funding originates in and is spent within economically rich countries (*4*). However, this leaves globally flexible funding of hundreds of millions of dollars annually from multilateral agencies (such as the Global Environment Facility), bilateral aid, and private sources including environmentally focused corporations, foundations, and individuals. These resources are frequently the only ones available where conservation is most needed, given that biodiversity is unevenly distributed and the most biodiverse places are often the most threatened and poorest economically (*5*). Accordingly, geographically flexible resources exert disproportionate influence on conservation worldwide and have a key role in the recently agreed-upon intergovernmental 2010 target to reduce significantly the rate of biodiversity loss (*6*).

The development of strategies to best allocate globally flexible conservation resources has attracted considerable attention since the pioneering work of Myers (*7*), resulting in much progress as well as much controversy. The wide variety of approaches has led to criticism that there is duplication of effort and lack of clarity (*8*). Although attempts have been made to summarize conservation planning strategies by scale (*9*), none has done so within the framework of conservation planning (*10*). We review the published concepts and methods behind global biodiversity conservation prioritization, assess the remaining challenges, and highlight how this synthesis can inform allocation of globally flexible resources.

Global Prioritization in Context

Nine major institutional templates of global biodiversity conservation prioritization have been published over the past decade, each with involvement from nongovernmental organizations (fig. S1). Conceptually, they all fit within the framework of "irreplaceability" relative to "vulnerability" (Fig. 1), which is central to conservation planning theory (*10*). However, they map onto different portions of the framework: Most of the templates prioritize high irreplaceability, but some prioritize high vulnerability and others prioritize low vulnerability. These differences are key to understanding how and why the nine prioritizations differ, yielding priority maps that cover from less than one-tenth to more than a third of Earth's land surface (Fig. 2).

Six of the nine templates of global conservation priority incorporate irreplaceability—measures of spatial conservation options (*10*). The most common measure of irreplaceability is plant (*11*–*14*) or bird (*15*) endemism, often supported by terrestrial vertebrate endemism overall (*11*, *13*, *14*). The logic for this is that greater the number of endemic species in a region, the more biodiversity is lost if that region is lost (although, in a strict sense, any location with even one endemic species is irreplaceable). In addition to the number of endemic species, other aspects of irreplaceability have been proposed, including taxonomic uniqueness, unusual phenomena, and global rarity of major habitat types (*16*), but these remain difficult to quantify. Although species richness within a given area is popularly assumed to be important in prioritization, none of the approaches relies on species richness alone. This is because species richness is driven by common, widespread species; thus, strategies focused on species richness tend to miss exactly those biodiversity features most in need of conservation (*17*, *18*). Three approaches do not incorporate irreplaceability (*19*–*21*).

The choice of irreplaceability measures is to some degree subjective, in that data limitations currently preclude the measurement of overall biodiversity. Furthermore, these data constraints mean that, with the exception of endemic bird areas (*15*), the measures of irreplaceability used in global conservation prioritization have been derived from the opinions of specialists. Subsequent tests of plant endemism estimates (*22*) have shown this expert opinion to be quite accurate. However, reliance on specialist opinion means that results cannot be replicated, raising questions concerning the transparency of the approaches (*8*). It also prevents a formal measurement of irreplaceability, which requires the identities of individual biodiversity features, such as species names, rather than just estimates of their magnitude expressed as a number (*8*, *23*).

Five of the templates of global conservation priority incorporate vulnerability—measures of temporal conservation options (*10*). A recent classification of vulnerability (*24*) recognizes four types of measures: (i) environmental and spatial variables, (ii) land tenure, (iii) threatened species, and (iv) expert opinion. Of these, environmental and spatial variables have been

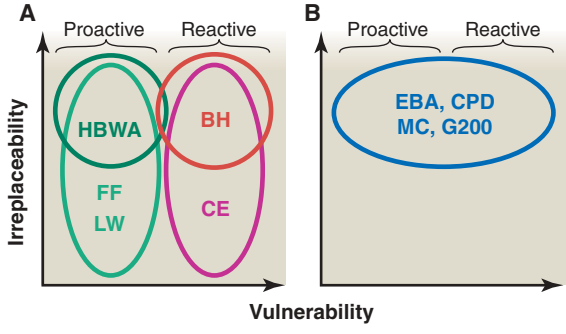


Fig. 1. Global biodiversity conservation priority templates placed within the conceptual framework of irreplaceability and vulnerability. Template names are spelled out in the Fig. 2 legend. (**A**) Purely reactive (prioritizing low vulnerability) and purely proactive (prioritizing high vulnerability) approaches. (**B**) Approaches that do not incorporate vulnerability as a criterion (all prioritize high irreplaceability).

[1]Conservation International, 1919 M Street, NW, Washington, DC 20036, USA. [2]World Agroforestry Centre (ICRAF), Post Office Box 35024, University of the Philippines, Los Baños, Laguna 4031, Philippines. [3]Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22904, USA. [4]Departamento de Zoologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270, Brazil. [5]Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK. [6]Nature Protection Trust of Seychelles, Post Office Box 207, Victoria, Mahé, Seychelles. [7]BirdLife International in Indochina, 4/209 Doi Can Street, Ba Dinh, Hanoi, Vietnam.

*To whom correspondence should be addressed. E-mail: t.brooks@conservation.org

used most frequently in global conservation prioritization, measured as proportionate habitat loss (*11*, *14*, *20*, *21*). Species-area relationships provide justification that habitat loss translates into biodiversity loss (*1*). However, the use of habitat loss as a measure of vulnerability has several problems: It is difficult to assess with the use of remote sensing for xeric and aquatic systems, it does not incorporate threats such as invasive species and hunting pressure, and it is retrospective rather than predictive (*24*). The frontier forests approach (*19*) uses absolute forest cover as a measure.

In addition to habitat loss, land tenure—measured as protected area coverage—has also been incorporated into two approaches (*16*, *21*). Other possible surrogates not classified by Wilson *et al*. (*24*) include human population growth and density, which are widely thought to be relevant (*25*–*27*) and were integral to two of the systems (*14*, *20*). None of the global conservation prioritization templates used threatened species or expert opinion as measures of vulnerability. Political and institutional capacity and governance (*27*) affect biodiversity indirectly, but have not been incorporated to date. This is true for climate change as well, which is of concern given that its impact is likely to be severe (*28*). Finally, although costs of conservation generally increase as the threat increases, no proposals for global biodiversity conservation priority have yet incorporated costs directly, despite the availability of techniques to do this at regional scales (*29*). Two of the templates of global conservation prioritization do not incorporate vulnerability (*12*, *13*), and the remaining two incorporate it only peripherally (*15*, *16*).

The spatial units most commonly used in systematic conservation planning are equal-area grids. However, data limitations have precluded their use in the development of actual templates of global biodiversity conservation priority to date. Instead, all proposals, with the exception of megadiversity countries (*13*), are based on biogeographic units. Typically, these units are defined a priori by specialist perception of the distribution of biodiversity. For example, "ecoregions," one of the most commonly used such classifications, are "relatively large units of land containing a characteristic set of natural communities that share a large majority of their species, dynamics, and environmental conditions" (*16*). Only in the endemic bird areas approach are biogeographic units defined a posteriori by the distributions of the species concerned (*15*). Relative to equal-area grids, biogeographic units bring advantages of ecological relevance, whereas megadiversity countries (*13*) bring political relevance.

Reliance on biogeographic spatial units raises several complications. Various competing bioregional classifications are in use (*30*), and the choice of system has considerable repercussions for resulting conservation priorities. Furthermore, when unequally sized units are used, priority may be biased toward large areas as a consequence of species-area relationships. Therefore, assessment of global conservation priorities should factor out area, either by taking residuals about a best-fit line to a plot of species against area (*18*) or by rescaling numbers of endemic species with the use of a power function (*23*). Nevertheless, the use of a priori bioregional units for global conservation prioritization will be essential until data of sufficient resolution become available to enable the use of grids.

In Fig. 3, we map the overlay of the global biodiversity conservation priority systems into geographic space from the conceptual framework of Fig. 1. Figure 3A illustrates the large degree of overlap between templates that prioritize highly vulnerable regions of high irreplaceability: tropical islands and mountains (including montane Mesoamerica, the Andes, the Brazilian Atlantic forest, Madagascar, montane Africa, the Western Ghats of India, Malaysia, Indonesia, the Philippines, and Hawaii), Mediterranean-type systems (including California, central Chile, coastal South Africa, southwest Australia, and the Mediterranean itself), and a few temperate forests (the Caucasus, the central Asian mountains, the Himalaya, and southwest China). Highly vulnerable regions of lower irreplaceability (generally, the rest of the northern temperate regions) are prioritized by fewer approaches. Figure 3B shows a large amount of overlap between templates for regions of low vulnerability but high irreplaceability, in particular the three major tropical rainforests of Amazonia, the Congo, and New Guinea. Regions of simultaneously lower vulnerability and irreplaceability, such as

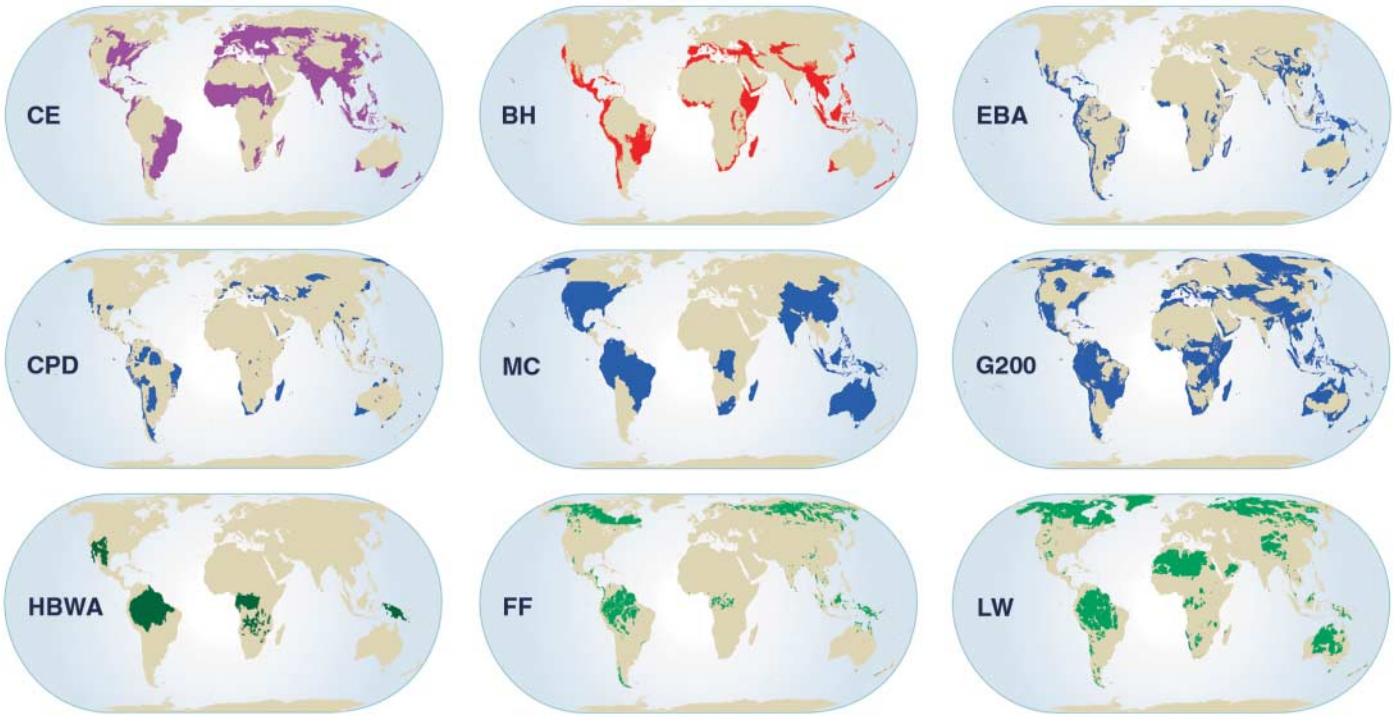


Fig. 2. Maps of the nine global biodiversity conservation priority templates: CE, crisis ecoregions (*21*); BH, biodiversity hot spots [(*11*), updated by (*39*)]; EBA, endemic bird areas (*15*); CPD, centers of plant diversity (*12*); MC, megadiversity countries (*13*); G200, global 200 ecoregions [(*16*), updated by (*54*)]; HBWA, high-biodiversity wilderness areas (*14*); FF, frontier forests (*19*); LW, last of the wild (*20*).

the boreal forests of Canada and Russia, and the deserts of the western United States and central Asia, are prioritized less often.

Two general observations are apparent. First, most land (79%) is highlighted by at least one of the prioritization systems. Second, despite this, a noticeable pattern emerges from the overlay of different approaches. There is significant overlap among templates that prioritize irreplaceable regions (*11*–*16*), among those that prioritize highly vulnerable regions (*11*, *21*), and among those that prioritize regions of low vulnerability (*14*, *19*, *20*), but not between approaches across each of these three general classes (table S1). This provides useful cross-verification of priority regions (*31*).

These patterns of overlap reflect two approaches to how vulnerability is incorporated into conservation in the broadest sense: reactive (prioritizing areas of high threat and high irreplaceability) and proactive (prioritizing areas of low threat but high irreplaceability). The former are considered the most urgent priorities in conservation planning theory (*10*) because unless immediate conservation action is taken within them, unique biodiversity will soon be lost. The latter are often de facto priorities, because the opportunities for conservation in these are considerable (*32*). Biodiversity conservation clearly needs both approaches, but the implementation of each may correspond to different methods. On the one hand, large-scale conservation initiatives may be possible in wilderness areas, such as the establishment of enormous protected areas (one example is the 3,800,000-ha Tumucumaque National Park, created in the Brazilian state of Amapá in 2003). On the other hand, finely tuned conservation will be essential in regions of simultaneously high irreplaceability and threat, where losing even tiny patches of remnant habitat, such as the sites identified by the Alliance for Zero Extinction (*33*), would be tragic.

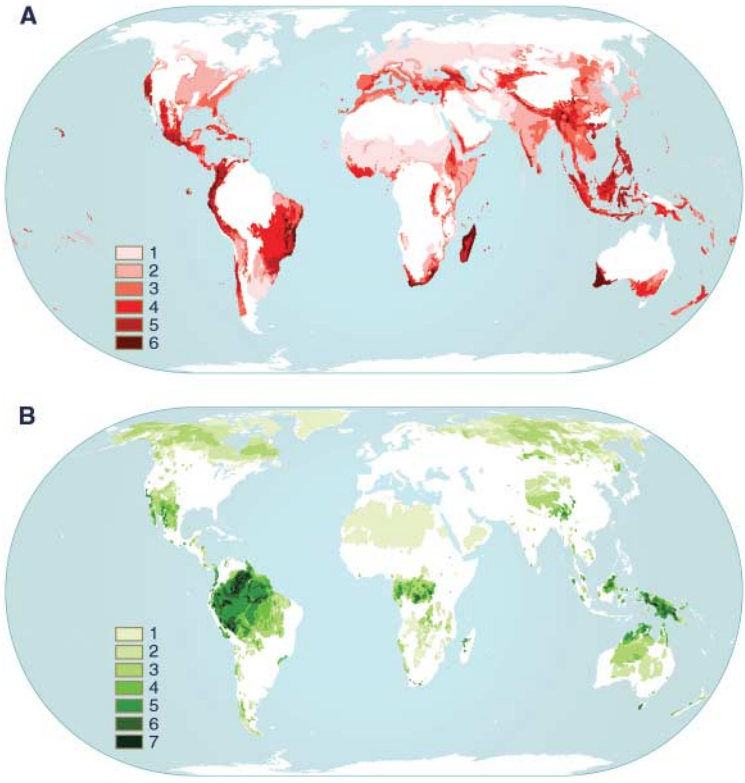


Fig. 3. Mapping the overlay of approaches prioritizing reactive and proactive conservation. (**A**) Reactive approaches, corresponding to the right-hand side of Fig. 1A, which prioritize regions of high threat, and those that do not incorporate vulnerability as a criterion (Fig. 1B); the latter are only mapped where they overlap with the former. (**B**) Proactive approaches, corresponding to the left-hand side of Fig. 1A, which prioritize regions of low threat, and those that do not incorporate vulnerability as a criterion (Fig. 1B); again, the latter are only mapped where they overlap with the former. Shading denotes the number of global biodiversity conservation prioritization templates that prioritize the shaded region, in both (A) and (B).

Impact of Global Prioritization

The appropriate measure of impact is the success of prioritization in achieving its main goal: influencing globally flexible donors to invest in regions where these funds can contribute most to conservation. Precise data are unavailable for all of the approaches (*34*), but hot spots alone have mobilized at least $750 million of funding for conservation in these regions (*35*). More specifically, conservation funding mechanisms have been established for several of the approaches, such as the $100 million, 10-year Global Conservation Fund focused on high-biodiversity wilderness areas and hot spots, and the $125 million, 5-year Critical Ecosystem Partnership Fund, aimed exclusively at hot spots. The Global Environment Facility, the largest financial mechanism addressing biodiversity conservation, is currently exploring a resource allocation framework that builds on existing templates. Both civil society and government organizations often use the recognition given to regions as global conservation priorities as justification when applying for geographically flexible funding. In addition, the global prioritization systems must have had sizeable effects in the cancellation, relocation, or mitigation of environmentally harmful activities, even in the absence of specific legislation. Unfortunately, resources still fall an order of magnitude short of required conservation funding (*4*). Nevertheless, the dollar amounts are impressive, and represent marked increases in conservation investment in these regions.

Challenges Facing Global Prioritization

Limitations of data have thus far generally restricted global conservation prioritization to specialist estimates of irreplaceability, to habitat loss as a measure of vulnerability, and to coarse geographic units defined a priori. Over the past 5 years, spatial data sets have been compiled with the potential to reduce these constraints, particularly for mammals, birds, and amphibians (*5*). When these maps are combined with assessment of conservation status, they enable the development of threat metrics directly based on threatened species (*36*). So far, the main advances to global prioritization enabled by these new data are validation tests of existing templates (*31*). Encouragingly, global gap analysis of priorities for the representation of terrestrial vertebrate species in protected areas (*36*) and initial regional assessment of plants (*37*) yield results similar to existing approaches (fig. S2).

Invertebrates represent the bulk of eukaryotic diversity on Earth with more than a million known species and many more yet to be described (*5*). The conservation status of only ~3500 arthropods has been assessed (*5*), so global conservation priority is far from being able to incorporate megadiverse invertebrate taxa (*8*, *23*). Although some regional data shows little overlap between priority areas for arthropods and those for plant and terrestrial vertebrate taxa (*38*), preliminary global data for groups such as tiger beetles and termites suggest much higher levels of congruence (*39*). Similarly, pioneering techniques to model overall irreplaceability by combining point data for megadiverse taxa with environmental data sets produce results commensurate with existing conservation priorities (*40*). These findings, although encouraging, in no way preclude the need to use primary invertebrate data in global conservation prioritization as they become available.

Aquatic systems feature poorly in existing conservation templates. Only one conservation prioritization explicitly incorporates aquatic systems (*16*). The most comprehensive study yet, albeit restricted to tropical coral reef ecosystems, identified 10 priority regions based on endemism and threat (*41*). Eight of these regions lie adjacent to priority regions highlighted in Fig. 3, raising the

possibility of correspondence between marine and terrestrial priorities, despite the expectation that surrogacy of conservation priorities will be low between different environments (*42*). Efforts to identify freshwater priorities lag further behind, although initial studies reveal a highly uneven distribution of freshwater fish endemism (*39*).

Most measurement of irreplaceability is species based, raising the concern that phylogenetic diversity may slip through the net of global conservation priorities (*8*, *23*, *43*). However, analyses for mammals (*44*) find that priority regions represent higher taxa and phylogenetic diversity better than would be predicted by the degree to which they represent species. Islands such as Madagascar and the Caribbean hold especially high concentrations of endemic genera and families (*39*). A heterodox perspective argues that the terminal tips of phylogenetic trees should be higher priorities than deep lineages (*45*). In any case, the balance of work implies that even if phylogenetic diversity is not explicitly targeted for conservation, global prioritization based on species provides a solid surrogate for evolutionary history.

That global conservation priority regions capture phylogenetic history does not necessarily mean that they represent evolutionary process (*8*). For example, transition zones or ''biogeographic crossroads,'' frequently overlooked by conservation prioritization, could be of particular importance in driving speciation (*46*). On the other hand, there is evidence that areas of greatest importance in generating biodiversity are those of long-term climatic stability, especially where they occur in tropical mountains (*47*), which are incorporated in most approaches to global conservation prioritization. The development of metrics for the maintenance of evolutionary process is in its infancy and represents an emerging research front.

A final dimension that will prove important to assess in the context of global conservation prioritization concerns ecosystem services (*43*). Although the processes threatening biodiversity and ecosystem services are likely similar, the relationship between biodiversity per se and ecosystem services remains unresolved (*48*). Thus, while it is important to establish distinct goals for these conservation objectives (*49*), identification of synergies between them is strategically vital. This research avenue has barely been explored, and questions of how global biodiversity conservation priorities overlap with priority regions for carbon sequestration, climate stabilization, maintenance of water quality, minimization of outbreaks of pests and diseases, and fisheries, for example, remain unanswered. However, the correspondence between conservation priorities and human populations (*25*, *26*) and poverty (*4*, *5*) is an indication that the conservation of areas of high biodiversity priority will deliver high local ecosystem service benefits.

From Global to Local Priorities

The establishment of global conservation priorities has been extremely influential in directing resources toward broad regions. However, a number of authors have pointed out that global conservation prioritization has had little success in informing actual conservation implementation (*8*, *23*). Separate processes are necessary to identify actual conservation targets and priorities at much finer scales, because even within a region as uniformly important as, for example, Madagascar, biodiversity and threats are not evenly distributed. Bottom-up processes of identification of priorities are therefore essential to ensure the implementation of area-based conservation (*50*).

Indeed, numerous efforts are underway to identify targets for conservation implementation. Many focus on the site scale, drawing on two decades of work across nearly 170 countries in the designation of important bird areas (*51*). There is an obvious need to expand such work to incorporate other taxa (*52*) and to prioritize the most threatened and irreplaceable sites (*33*). Such initiatives have recently gained strong political support under the Convention on Biological Diversity, through the development of the Global Strategy for Plant Conservation and the Programme of Work on Protected Areas. Both mechanisms call for the identification, recognition, and safeguarding of sites of biodiversity conservation importance. Meanwhile, considerable attention is also targeted at the scale of landscapes and seascapes to ensure not just the representation of biodiversity but also of the connectivity, spatial structure, and processes that allow its persistence (*53*).

Global conservation planning is key for strategic allocation of flexible resources. Despite divergence in methods between the different schemes, an overall picture is emerging in which a few regions, particularly in the tropics and in Mediterranean-type environments, are consistently emphasized as priorities for biodiversity conservation. It is crucial that the global donor community channel sufficient resources to these regions, at the very minimum. This focus will continue to improve if the rigor and breadth of biodiversity and threat data continue to be consolidated, which is especially important given the increased accountability demanded from global donors. However, it is through the conservation of actual sites that biodiversity will ultimately be preserved or lost, and thus drawing the lessons of global conservation prioritization down to a much finer scale is now the primary concern for conservation planning.

References and Notes

1. S. L. Pimm, G. J. Russell, J. L. Gittleman, T. M. Brooks, *Science* **269**, 347 (1995).
2. Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being* (Island Press, Washington, DC, 2005).
3. M. L. Hunter Jr., A. Hutchinson, *Conserv. Biol.* **8**, 1163 (1994).
4. A. James, K. J. Gaston, A. Balmford, *Nature* **401**, 323 (1999).
5. J. E. M. Baillie *et al.*, *Global Species Assessment* (IUCN, Gland, Switzerland, 2004).
6. A. Balmford *et al.*, *Science* **307**, 212 (2005).
7. N. Myers, *Environmentalist* **8**, 187 (1988).
8. G. M. Mace *et al.*, *Nature* **405**, 393 (2000).
9. K. H. Redford *et al.*, *Conserv. Biol.* **17**, 116 (2003).
10. C. R. Margules, R. L. Pressey, *Nature* **405**, 243 (2000).
11. N. Myers *et al.*, *Nature* **403**, 853 (2000).
12. WWF, IUCN, *Centres of Plant Diversity* (WWF and IUCN, Gland, Switzerland, 1994–1997).
13. R. A. Mittermeier, P. Robles Gil, C. G. Mittermeier, *Megadiversity* (CEMEX, Mexico City, Mexico, 1997).
14. R. A. Mittermeier *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 10309 (2003).
15. A. J. Stattersfield, M. J. Crosby, A. J. Long, D. C. Wege, *Endemic Bird Areas of the World* (BirdLife International, Cambridge, UK, 1998).
16. D. M. Olson, E. Dinerstein, *Conserv. Biol.* **12**, 502 (1998).
17. C. D. L. Orme *et al.*, *Nature* **436**, 1016 (2005).
18. J. F. Lamoreux *et al.*, *Nature* **440**, 212 (2006).
19. D. Bryant, D. Nielsen, L. Tangley, *Last Frontier Forests* (World Resources Institute, Washington, DC, 1997).
20. E. W. Sanderson *et al.*, *Bioscience* **52**, 891 (2002).
21. J. M. Hoekstra *et al.*, *Ecol. Lett.* **8**, 23 (2005).
22. G. A. Krupnick, W. J. Kress, *Biodivers. Conserv.* **12**, 2237 (2003).
23. N. Brummitt, E. N. Lughadha, *Conserv. Biol.* **17**, 1442 (2003).
24. K. Wilson *et al.*, *Environ. Manage.* **35**, 527 (2005).
25. R. P. Cincotta, J. Wisnewski, R. Engelman, *Nature* **404**, 990 (2000).
26. A. Balmford *et al.*, *Science* **291**, 2616 (2001).
27. C. O'Connor, M. Marvier, P. Kareiva, *Ecol. Lett.* **6**, 706 (2003).
28. C. Parmesan, G. Yohe, *Nature* **421**, 37 (2003).
29. K. A. Wilson *et al.*, *Nature* **440**, 337 (2006).
30. P. Jepson, R. J. Whittaker, *Conserv. Biol.* **16**, 42 (2002).
31. G. A. B. da Fonseca *et al.*, *Nature* **405**, 393 (2000).
32. M. Cardillo, G. M. Mace, J. L. Gittleman, A. Purvis, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 4157 (2006).
33. T. H. Ricketts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 18497 (2005).
34. B. S. Halpern *et al.*, *Conserv. Biol.* **20**, 56 (2006).
35. N. Myers, *Bioscience* **53**, 916 (2003).
36. A. S. L. Rodrigues *et al.*, *Bioscience* **54**, 1092 (2004).
37. W. Küper *et al.*, *Ann. Missouri Bot. Gard.* **91**, 525 (2004).
38. A. P. Dobson, J. P. Rodriguez, W. M. Roberts, D. S. Wilcove, *Science* **275**, 550 (1997).
39. R. A. Mittermeier *et al.*, *Hotspots Revisited* (CEMEX, Mexico City, Mexico, 2004).
40. S. Ferrier *et al.*, *Bioscience* **54**, 1101 (2004).
41. C. M. Roberts *et al.*, *Science* **295**, 1280 (2002).
42. W. V. Reid, *Trends Ecol. Evol.* **13**, 275 (1998).
43. P. Kareiva, M. Marvier, *Am. Sci.* **91**, 344 (2003).
44. W. Sechrest *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2067 (2002).
45. T. L. Erwin, *Science* **253**, 750 (1991).
46. T. B. Smith, R. K. Wayne, D. J. Girman, M. W. Bruford, *Science* **276**, 1855 (1997).
47. J. Fjeldså, J. C. Lovett, *Biodivers. Conserv.* **6**, 325 (1997).
48. M. Loreau *et al.*, *Science* **294**, 804 (2001).
49. S. Sarkar, *Bioscience* **49**, 405 (1999).
50. R. J. Whittaker *et al.*, *Divers. Distrib.* **11**, 3 (2005).
51. BirdLife International, *State of the World's Birds 2004* (BirdLife International, Cambridge, UK, 2004).
52. G. Eken *et al.*, *Bioscience* **54**, 1110 (2004).
53. R. M. Cowling, R. L. Pressey, M. Rouget, A. T. Lombard, *Biol. Conserv.* **112**, 191 (2003).
54. D. M. Olson, E. Dinerstein, *Ann. Missouri Bot. Gard.* **89**, 199 (2002).

55. We thank G. Fabregas, D. Knox, T. Lacher, P. Langhammer, N. Myers, and W. Turner for help with the manuscript, and the Gordon and Betty Moore Foundation for funding.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/58/DC1
Figs. S1 and S2
Table S1
References and Notes

10.1126/science.1127609

Nuclear Gene Indicates Coat-Color Polymorphism in Mammoths

Holger Römpler,[1] Nadin Rohland,[2] Carles Lalueza-Fox,[3] Eske Willerslev,[4] Tatyana Kuznetsova,[5] Gernot Rabeder,[6] Jaume Bertranpetit,[7] Torsten Schöneberg,[1] Michael Hofreiter[2]*

Because more than 99% of all species that have ever lived on Earth are extinct, the genetic basis of most phenotypic traits that have evolved during life's history cannot be studied directly. Although it is possible to approximately reconstruct the gene sequences of extant species' common ancestors (*1*), many extinct species represent end points of evolution, and their gene sequences cannot be investigated by using this approach. For late Pleistocene species, ancient DNA techniques should in principle allow analyses of extinct genes. However, despite progress in the analysis of ancient DNA, to date only short fragments of nuclear genes could be amplified from Pleistocene specimens (*2*, *3*).

We have combined two methods recently introduced in ancient DNA research, large-scale extraction and multiplex polymerase chain reaction (PCR), to amplify 1236 base pairs (bp), including the complete coding sequence, of the single-exon nuclear melanocortin type 1 receptor (*Mc1r*) gene from an ~43,000-year-old mammoth (*Mammuthus primigenius*) bone from Siberia. The product of the *Mc1r* gene is a key protein in the determination of hair color in both humans and other animals (*4*), providing an opportunity to connect genotype and phenotype in extinct taxa. For example, reduced activity of MC1R has been shown to result in red hair in humans and cows and red or yellow hair in mice, horses, and dogs (*4*). Dark- and light-colored mammoth hair has been recovered from permafrost mummies, giving circumstantial evidence that mutations may occur in one of the genes involved in determination of hair color. We amplified the mammoth *Mc1r* gene in 23 primary fragments varying in length from 101 to 138 bp. Three positions were found to be polymorphic in the individual sequenced, all resulting in amino acid changes [Thr^{21}→Ala^{21} (Thr^{21}Ala), Arg^{67}Cys, and Arg^{301}Ser; positions relative to the elephant *Mc1r* sequence] (Fig. 1A) (*5*). Because template damage may affect ancient DNA sequences (*6*), we sequenced each polymorphic position from at least three independent PCRs and, by using single-nucleotide polymorphism (SNP) typing, from additional specimens. The identification of homozygous individuals from these additional mammoth samples revealed most likely allelic constitutions of Thr^{21}-Arg^{67}-Arg^{301} (allele 1) and Ala^{21}-Cys^{67}-Ser^{301} (allele 2), respectively (*5*). Whereas Thr^{21} is highly variable among mammalian MC1R orthologs, substitutions of Arg^{67} and Arg^{301} (Fig. 1A) affect highly conserved residues. To investigate whether the two alleles show a functional difference, we constructed *Mc1r* expression vectors that are identical to the inferred mammoth alleles on the amino acid level. The product of allele 1 shows high basal activity and robust responses to two agonists, αMSH (Fig. 1B) and NDP-αMSH (*5*), in a concentration-dependent manner. In contrast, the product of allele 2 shows an almost complete loss in basal activity and an ~65% reduction in efficacy (E_{max}) to both agonists. Functional testing of each position variable in the mammoth individually and all possible combinations of the three mammoth-specific amino acids revealed the Arg^{67}Cys substitution as the main cause for MC1R dysfunction (Fig. 1B) (*5*).

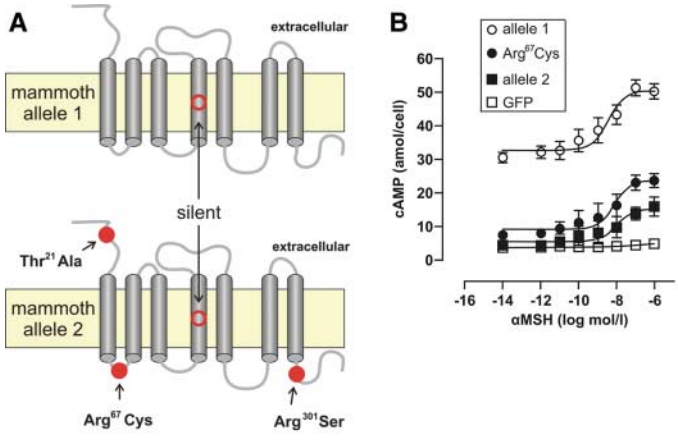


Fig. 1. (**A**) Localization of the fixed and polymorphic positions in mammoth *Mc1r*. The seven transmembrane domain architecture of MC1R is depicted. Compared with the *Mc1r* sequence of the African elephant, the approximate positions of the fixed silent position and polymorphic positions in the two mammoth alleles are circled. (**B**) Partial loss of function in mammoth MC1R variants. For functional characterization of the wild type (allele 1), Arg^{67}Cys and Thr^{21}Ala–Arg^{67}Cys–Arg^{301}Ser (allele 2) constructs of the *Mc1r* gene were transiently transfected into COS-7 cells and tested for agonist-induced cyclic adenosine monophosphate (cAMP) accumulation. Green fluorescent protein (GFP) plasmid–transfected cells served as controls. Intracellular cAMP was measured in response to increasing concentrations of α-MSH as described (*5*). Shown is the mean ± SEM of four independent experiments, each performed in triplicate.

Pleistocene mammoths harbored both a fully and a partially active MC1R variant. Given an otherwise functional pathway for hair coloration, the reduction in activity of the Arg^{67}Cys variant would be sufficient to result in substantially lighter hair color. Intriguingly, light-colored populations of the beach mouse (*Peromyscus polionotus leucocephalus*) carry an *Mc1r* allele that contains an Arg-to-Cys substitution at the homologous sequence position, and functional tests and crossing experiments have revealed both a reduction in basal and induced activity highly similar to that observed for the mammoth MC1R protein and a strong association between this amino acid polymorphism and adaptive coat color phenotype (*7*). These findings strongly argue for a phenotypic effect of this mutation also in the mammoth. Thus, it is possible that mammoth populations were polymorphic with regard to hair color, harboring both dark- and light-haired individuals. Although it is currently impossible to determine whether hair color was adaptive in mammoth, our study shows that functional analyses of nuclear genes are possible from late Pleistocene species. This opens up the possibility of studying a wide range of extinct species' features invisible in the fossil record.

References and Notes

1. J. W. Thornton, *Nat. Rev. Genet.* **5**, 366 (2004).
2. A. Greenwood, C. Capelli, G. Possnert, S. Pääbo, *Mol. Biol. Evol.* **16**, 1466 (1999).
3. H. N. Poinar *et al.*, *Science* **311**, 392 (2006).
4. J. L. Rees, *Annu. Rev. Genet.* **37**, 67 (2003).
5. Materials and methods are available as supporting material on *Science* Online.
6. M. Hofreiter *et al.*, *Nucleic Acids Res.* **29**, 4793 (2001).
7. H. E. Hoekstra, R. J. Hirschmann, R. A. Bundey, P. A. Insel, J. P. Crossland, *Science* **313**, 101 (2006).
8. We thank the members of our research groups and several anonymous reviewers for comments that improved the manuscript. This work was funded by the Max Planck Gesellschaft, Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie and Deutsche Forschungsgemeinschaft. Sequences have been deposited in GenBank under accession nos. DQ648859 to DQ648866.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/62/DC1
Materials and Methods
Tables S1 and S2
References

20 April 2006; accepted 30 May 2006
10.1126/science.1128994

[1]Molecular Biochemistry, Institute of Biochemistry, Medical Faculty, University of Leipzig, 04103 Leipzig, Germany. [2]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. [3]Departament de Biologia Animal, Facultat de Biologia, Universitat de Barcelona, 08028 Barcelona, Spain. [4]Ancient DNA and Evolution Group, Centre for Ancient Genetics, Niels Bohr Institute and Biological Institute, University of Copenhagen, 2100 Copenhagen, Denmark. [5]Department of Paleontology, Faculty of Geology, Moscow State University, 119992 Moscow, Russia. [6]Department of Paleontology, University of Vienna, 1090 Vienna, Austria. [7]Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, 08003 Barcelona, Spain.

*To whom correspondence should be addressed. E-mail: hofreite@eva.mpg.de

Hox Control of Organ Size by Regulation of Morphogen Production and Mobility

Michael A. Crickmore[1] and Richard S. Mann[2]*

Selector genes modify developmental pathways to sculpt animal body parts. Although body parts differ in size, the ways in which selector genes create size differences are unknown. We have studied how the *Drosophila* Hox gene *Ultrabithorax* (*Ubx*) limits the size of the haltere, which, by the end of larval development, has ~fivefold fewer cells than the wing. We find that *Ubx* controls haltere size by restricting both the transcription and the mobility of the morphogen Decapentaplegic (Dpp). *Ubx* restricts Dpp's distribution in the haltere by increasing the amounts of the Dpp receptor, *thickveins*. Because morphogens control tissue growth in many contexts, these findings provide a potentially general mechanism for how selector genes modify organ sizes.

Changes in body part sizes have been critical for diversification and specialization of animal species during evolution. The beaks of Darwin's finches provide a famous example for how adaptation can produce variations in size and shape that allowed these birds to take advantage of specialized ecological niches and food supplies (*1*). Sizes also vary between homologous structures within an individual. For example, vertebrate digits and ribs vary in size, likely due to the activities of selector genes such as the Hox genes (*2*–*4*). Although the control of organ growth by selector genes is likely to be common in animal development (*2*, *5*, *6*), little is known about the mechanisms underlying this control.

The two flight appendages of *Drosophila melanogaster*, the wing and the haltere, provide a classic example of serially homologous structures of different sizes (Fig. 1A). Halteres, appendages used for balance during flight, are thought to have been modified from full-sized hindwings during the evolution of two-winged flies from their four-winged ancestors (*7*, *8*). All aspects of haltere development that distinguish it from a wing, including its reduced size, are under the control of the Hox gene *Ultrabithorax* (*Ubx*), which is expressed in all haltere imaginal disc cells but not in wing imaginal disc cells (*8*, *9*) (Fig. 1B). At all stages of development, haltere and wing primordia (imaginal discs) are different sizes. In the embryo, the wing primordium has about twice as many cells as the haltere primordium (*7*, *10*). By the end of larval development, the wing disc has ~five times more cells (~50,000) than the haltere disc (~10,000) (*11*) [Fig. 1B and Supporting Online Material (SOM) Text]. The wing and haltere appendages will form from the pouch region of these mature discs (fig. S1). The final step that contributes to wing and haltere size differences occurs during metamorphosis, when

[1]Department of Biological Sciences, Columbia University, New York, NY 10027, USA. [2]Department of Biochemistry and Molecular Biophysics, Columbia University, HHSC 1104, 701 West 168th Street, New York, NY 10032, USA.

*To whom correspondence should be addressed. E-mail: rsm10@columbia.edu

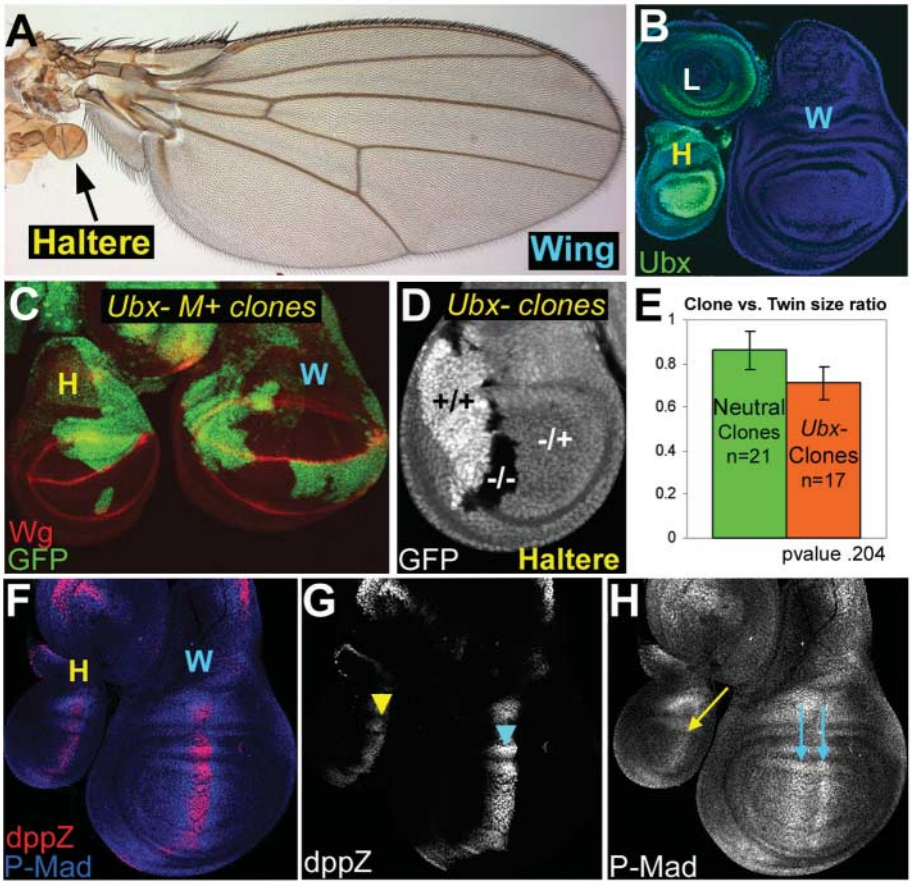


Fig. 1. Reduced Dpp production and transduction in the haltere. (**A**) Wild-type adult wing and haltere (arrow). (**B**) Third instar wing (W), haltere (H), and T3 leg (L) imaginal discs stained for Ubx (green) and a ubiquitous nuclear protein (blue). Ubx is present in all haltere disc cells but not in wing disc cells. (**C**) Removing *Ubx* activity (lack of GFP) from more than 50% of a haltere disc during larval development using the M^+ (*Minute*) technique (*13*) (SOM Text) increased its size [compare with discs in (B) and (F)]. (**D**) Isolated Ubx^- clones (black, $-/-$) were not larger than Ubx^+ twin spots (bright white, $+/+$) in a *Ubx* heterozygous haltere (gray, $-/+$). (**E**) *Ubx* mutant:twin spot and neutral:twin spot clone size ratios. Error bars indicate SEM. (**F** to **H**) Wild-type wing and haltere discs stained for *dpp-lacZ* and P-Mad patterns. In the haltere, *dpp-lacZ* was reduced (arrowheads) and overlapped with a compacted P-Mad gradient (arrows).

wing, but not haltere, cells flatten, thus increasing the surface area of the final appendage (*12*).

Nonautonomous control of haltere size by *Ubx*. To confirm that *Ubx* has a postembryonic role in limiting the size of the haltere disc, we generated Ubx^- clones midway through larval development (*13*). Haltere discs–bearing large Ubx^- clones generated at this time become much larger than wild-type discs (Fig. 1C and SOM Text). *Ubx* could limit haltere size cell-autonomously by, for example, slowing the cell cycle of haltere cells relative to wing cells. We tested this by comparing the sizes of isolated Ubx^- clones in the haltere with those of their simultaneously generated wild-type twin clones. Contrary to the prediction of a cell-autonomous function for *Ubx* in size control, *Ubx* mutant clones did not grow larger than their twins (Fig. 1, D and E), a result that is consistent with earlier experiments suggesting that wing and haltere cells have similar mitotic rates during development (*14*). Hence, *Ubx* limits the size of the haltere during larval development by modifying pathways that control organ growth cell-nonautonomously.

***Ubx* regulation of Dpp signaling.** In the fly wing, Decapentaplegic (Dpp) [a long-range morphogen of the bone morphogenetic protein (BMP) family] has been shown to promote growth (*15*–*17*). In both the wing and the haltere, Dpp is produced and secreted from a specialized stripe of cells called the AP organizer, which is induced by the juxtaposition of anterior (A) and posterior (P) compartments, two groups of cells that have separate cell lineages (*18*). The AP organizer is a stripe of A cells that are instructed to synthesize Dpp by the short-range morphogen Hedgehog (Hh) secreted from adjacent P compartment cells (*18*–*22*). Dpp has a positive role in appendage growth. When more Dpp is supplied to the wing disc, either ectopically or within the AP organizer, more cells are incorporated into the developing wing field (*22*–*24*). Conversely, mutations that reduce the amount of Dpp lead to smaller wings (fig. S3) (*25*).

A comparison of the expression patterns of Dpp pathway components in the wing and the haltere demonstrates that *Ubx* is modifying this pathway (Fig. 1, F to H, fig. S1, and SOM Text). Compared with the wing, the stripe of *dpp* expression in the haltere was reduced in both its width and intensity, as reported by a *lacZ* insertion into the *dpp* locus (*dpp-lacZ*). There was also a difference in the profile of Dpp pathway activation, as visualized by an antibody that detects P-Mad, the activated form of the Dpp pathway transcription factor Mothers against Dpp (Mad). In the wing, P-Mad staining was low in the cells that transcribe *dpp* (Fig. 1, F to H, and fig. S1) (*26*). Immediately anterior and posterior to this activity trough, P-Mad labeling peaked in intensity and then gradually decayed further from the Dpp source, revealing a bimodal activity gradient. In contrast, in the haltere intense P-Mad staining was detected only in a single stripe of cells that overlaps with Dpp-producing cells of the AP organizer (Fig. 1, F to H, and fig. S1).

Because of the coincidence between *dpp* transcription and peak P-Mad staining in the haltere, we hypothesized that Dpp might be less able to move from haltere cells that secrete this ligand. We tested this idea by generating clones of cells in both wing and haltere discs in which the *actin5c* promoter drove the expression of a green fluorescent protein (GFP)–tagged version of Dpp (Dpp:GFP) (*13*, *27*, *28*). By using an extracellular staining protocol to analyze simultaneously generated clones (*29*), we observed Dpp:GFP and P-Mad much further from producing cells in the wing than in the haltere (Fig. 2, A to D). These observations strongly suggest that, compared with the wing, Dpp's mobility—and consequently the range of Dpp pathway activation—is reduced in the haltere.

We also tested whether the decreased production of Dpp in the haltere contributes to the different pattern of pathway activation observed in this tissue compared with the wing. This is unlikely because, even in haltere discs that overexpress Dpp in its normal expression domain, peak P-Mad staining was still observed close to Dpp-expressing cells (Fig. 2, E and F) (*13*). Despite increased *dpp* expression, no P-Mad activity trough was observed in these haltere discs. Further, although they become larger, these discs remained smaller than wild-type wing discs. We conclude that the decreased Dpp production in the haltere contributes to its reduced growth, but there must be mechanisms that also limit the extent of Dpp pathway activation, even in the presence of increased Dpp production.

One way in which Dpp's activation profile can be modified is by varying the production of the type I Dpp receptor, Thickveins (Tkv) (*26*, *30*). In the wing, *tkv* expression is low within and around the source of Dpp, resulting in low Dpp signal transduction in these cells and robust Dpp diffusion (*26*, *30*, *31*) (Fig. 3, A and B, and fig. S1). Low *tkv* expression in the medial wing is due to repression by both Hh and Dpp (*26*, *30*). Accordingly, *tkv* expression is highest in lateral regions of the wing disc, where Hh and Dpp signaling are low. In contrast to the wing, *tkv* transcription and protein levels were high in all cells of the haltere (Fig. 3, A and B). Thus, the more restricted Dpp mobility and P-Mad pattern in the haltere may result from a failure to repress *tkv* medially. To test this idea, we supplied all cells of the wing disc with uniform $UAS\text{-}tkv^+$ expression, to mimic the haltere pattern (Fig.

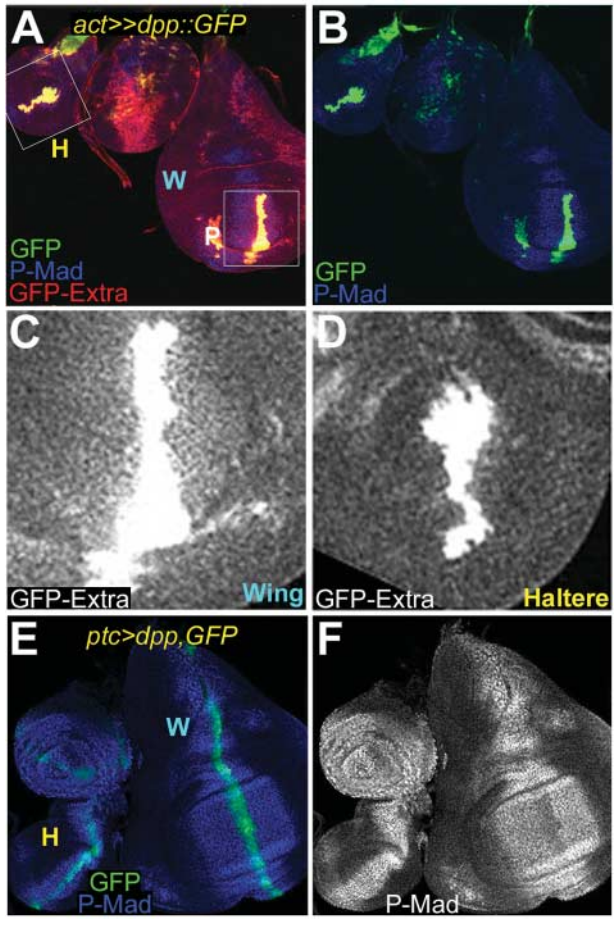


Fig. 2. Reduced Dpp mobility in the haltere. (**A** to **D**) Simultaneously generated *actin5c* promoter flp-out clones expressing *UAS-dpp::gfp* and *UAS-GFP* in the wing and haltere stained for extracellular GFP (red and white) and P-Mad (blue). The green channel shows GFP autofluorescence and marks the clone. The extracellular Dpp::GFP pattern closely correlates with the P-Mad pattern. In (C) and (D), enlarged images of the regions boxed in (A) are shown. (**E** and **F**) Overexpressing *dpp* with *ptc-Gal4* (visualized with *UAS-GFP*) increased the scale and intensity of P-Mad staining in the wing and the haltere, but the patterns remained qualitatively similar to those of wild type.

3C) (*13*). The resulting P-Mad pattern in these wing discs was very similar to that found in the wild-type haltere: The P-Mad trough was gone, and the activity gradient was compacted into a single stripe that coincides with Dpp-producing cells. Conversely, lowering the amount of Tkv in the haltere by expressing an RNA interference (RNAi) hairpin construct directed against *tkv* (*UAS-tkvRNAi*) in Dpp-producing cells induced a bimodal pattern of P-Mad staining similar to that of the wild-type wing disc (Fig. 3, D to F) (*13*). Thus, different amounts of Tkv result in qualitative differences in the P-Mad profiles of the wing and the haltere.

***tkv* expression and appendage size.** We hypothesized that the more limited pathway activation in the haltere might contribute to its smaller size. If correct, increasing *tkv* expression in the wing should reduce its size. Adult wings from flies expressing uniform *UAS-tkv*$^{+}$ were ~30% smaller than control wings; however, wing cell size remained the same (Fig. 3, G and H, and fig. S2) (*13*, *30*). Similar results were seen in staged imaginal discs and when *UAS-tkv*$^{+}$ expression was limited to the wing and the haltere (fig. S2). Conversely, reducing Tkv amounts by uniformly expressing *UAS-tkvRNAi* in wings and halteres increased haltere size by 30 to 60% (Fig. 3, I and J). In a complementary experiment, we reduced *tkv* transcription in the haltere by expressing a known *tkv* repressor, *master of thickveins* (*mtv*) (*32*). In this experiment, we measured haltere discs instead of the adult appendage and found, consistently, that the appendage-generating region of these discs increased in size by ~40% (fig. S2). Thus, different amounts of Tkv not only affect Dpp pathway activation but also affect organ size. The fact that manipulating only Tkv production does not fully transform the sizes of these appendages suggests that additional mechanisms, such as the reduced amounts of *dpp* transcription and the modulation of other morphogen pathways by *Ubx*, also contribute to size regulation. Consistently, when Dpp production is decreased in wing discs that uniformly express *UAS-tkv*$^{+}$, wing size was reduced more than it was by either single manipulation (fig. S3).

***Ubx* regulation of *tkv*.** We next address how *Ubx* up-regulates *tkv* in the haltere. *tkv-lacZ* expression and amounts of Tkv protein were cell-autonomously reduced in medial *Ubx*$^{-}$ clones, whereas lateral *Ubx* mutant tissue retained high amounts of Tkv (Fig. 4, A to D, and fig. S4). Because *tkv* is repressed by Dpp and Hh signaling in the wing (*26*, *30*), these results suggest that, in the haltere, these signals are not able to repress *tkv*. Consistently, activation of the Dpp pathway by expressing a constitutively active form of Tkv (Tkv^{QD}) resulted in cell-autonomous *tkv-lacZ* repression in the wing pouch (Fig. 4, E and F), whereas repression is not observed in the corresponding region of the haltere disc (Fig. 4, G and H).

In *Ubx* mosaic haltere discs, we also found that medial *Ubx*$^{+}$ tissue showed stronger P-Mad staining than *Ubx*$^{-}$ tissue at the same distance from the Dpp source (Fig. 4, A to D). We interpret this observation as evidence that *Ubx*$^{+}$ tissue is more effective at trapping and transducing Dpp than *Ubx*$^{-}$ tissue because of higher Tkv production in *Ubx*$^{+}$ cells.

To further understand the control of *tkv* by *Ubx*, we examined the known *tkv* repressor, *mtv* (*32*). In medial wing disc cells, *mtv* expression is approximately complementary to *tkv* expression (Fig. 4, I and J, and fig. S1), and *mtv*$^{-}$ clones in this region of the wing disc cell autonomously derepressed *tkv* (fig. S4) (*32*). In the haltere, very low *mtv-lacZ* expression was detected in the cells that stained strongly for P-Mad, suggesting that *mtv* is repressed by Dpp in this appendage (Fig. 4, I and J). Accordingly, strong repression of *mtv-lacZ* was seen in *UAS-tkv*^{QD}-expressing haltere pouch clones, whereas weak or no repression was seen in analogous wing clones (Fig. 4, K and L). We also found that, as expected, *Ubx*$^{-}$ clones in the medial haltere cell autonomously derepressed *mtv-lacZ* (fig. S4).

In the wing, Dpp and *mtv* are mandatory repressors of *tkv*: In the absence of either, *tkv* expression is high. In the haltere in the presence of *Ubx*, Dpp is a repressor of *mtv*. Consequently, high levels of these obligate *tkv* repressors (Dpp signaling and *mtv*) do not coexist in the haltere, resulting in *tkv* derepression. Consistent with this model, when we forced *mtv* expression in the medial haltere, where it coexists with Dpp signaling, it repressed *tkv-lacZ* (fig. S4). We note, however, that *Ubx* is likely to control *tkv* through additional means, because *mtv* mutant wing clones did not derepress *tkv-lacZ* expression to

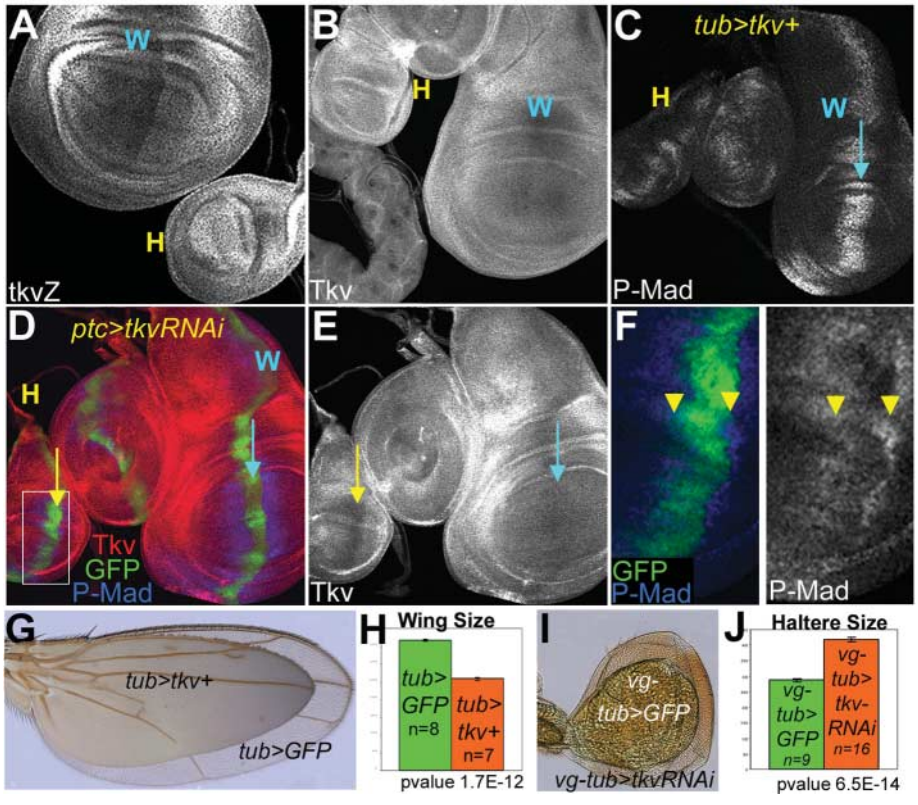


Fig. 3. Tkv production influences Dpp signaling and appendage size. (**A**) *tkv-lacZ* expression was high throughout the haltere, whereas in the wing it was low medially and high laterally. (**B**) Tkv antibody staining showed a pattern similar to that of the *tkv-lacZ* enhancer trap, with less resolution. (**C**) Driving uniform *UAS-tkv*$^{+}$ expression in the wing using *tubGal4* compacted the Dpp activity gradient and created haltere-like P-Mad staining pattern (arrow). (**D** to **F**) Expressing *UAS-tkvRNAi* in the haltere using *ptc-Gal4* (visualized with *UAS-GFP*) reduced Tkv staining [yellow arrow in (D) and (E)] and resulted in a bimodal P-Mad staining pattern [yellow arrowheads in (F), which shows a magnification of the region boxed in (D)]. (**G**) Adult wings uniformly expressing *UAS-tkv*$^{+}$ using *tubGal4* were ~30% smaller than control wings. (**H**) Quantification of wing size reduction caused by uniform *UAS-tkv*$^{+}$ expression (orange) compared to controls (green). Error bars indicate SEM. (**I**) Adult halteres uniformly expressing *UAS-tkvRNAi* using *vg-tubGal4* were up to 60% larger than control halteres. (**J**) Quantification of haltere size increase caused by uniform *tkvRNAi* expression (orange) compared to controls (green). The average increase seen is 46%.

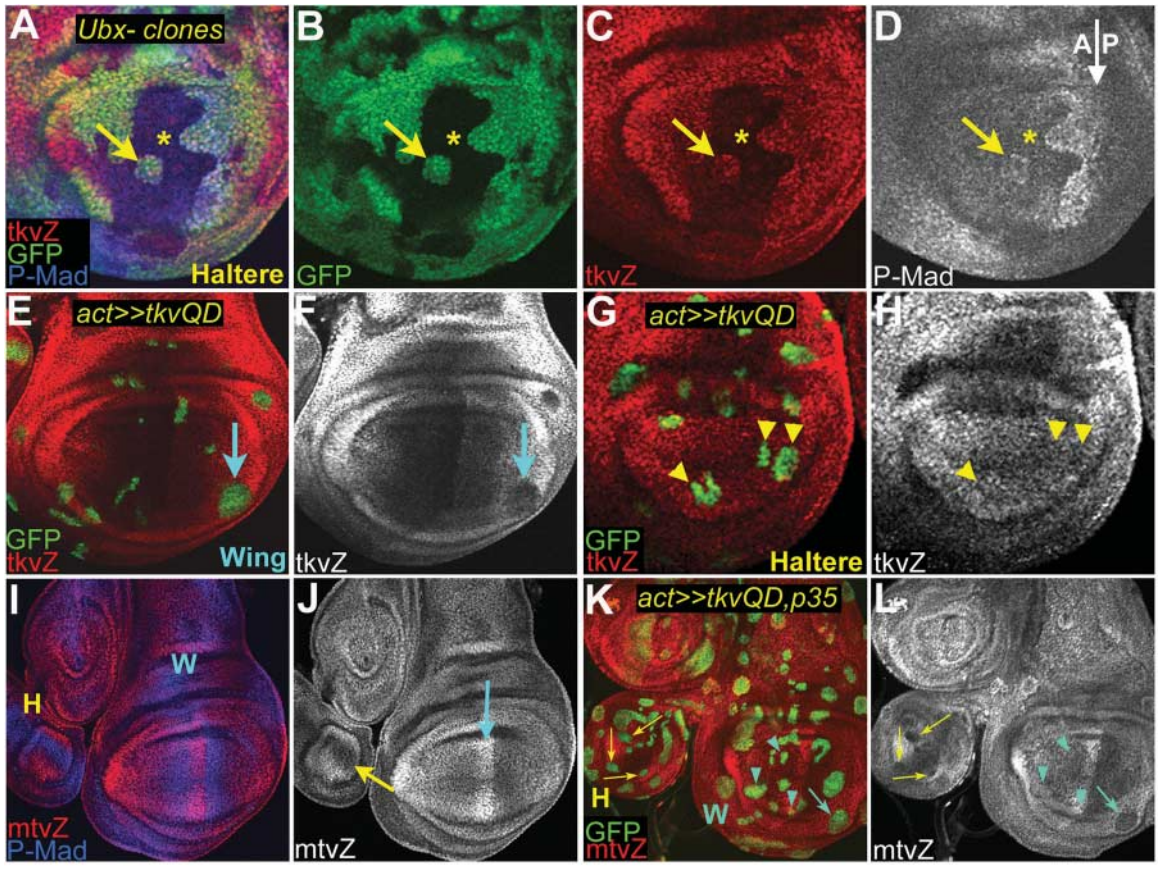


Fig. 4. Dpp and *Ubx* collaborate to repress a *tkv* repressor in the haltere. (**A** to **D**) *Ubx* mutant tissue in the medial haltere (absence of GFP) shows a cell-autonomous reduction in *tkv-lacZ* and reduced P-Mad staining. High P-Mad and *tkv-lacZ* staining can be detected in a Ubx^{+} island (yellow arrow) that is separated from Dpp-producing cells by Ubx^{-} tissue (*). The approximate position of the AP boundary is indicated by a white arrow in (D). (**E** to **H**) Clones expressing *UAS-tkv*^{QD} (marked with GFP) repress *tkv-lacZ* in the wing pouch (cyan arrow) but not in the analogous domain of the haltere (yellow arrowheads). (**I** and **J**) Wild-type wing and haltere discs stained for *mtv-lacZ* and P-Mad. *mtv-lacZ* is strongly expressed in Dpp-producing cells of the wing (cyan arrow) but is repressed in Dpp-producing cells of the haltere (yellow arrow). (**K** and **L**) Clones expressing *UAS-tkv*^{QD} (marked with GFP) strongly repress *mtv-lacZ* in the haltere (yellow arrows). Similar clones in the wing repress *mtv-lacZ* moderately in the P compartment (cyan arrow) and not at all in the A compartment (cyan arrowheads).

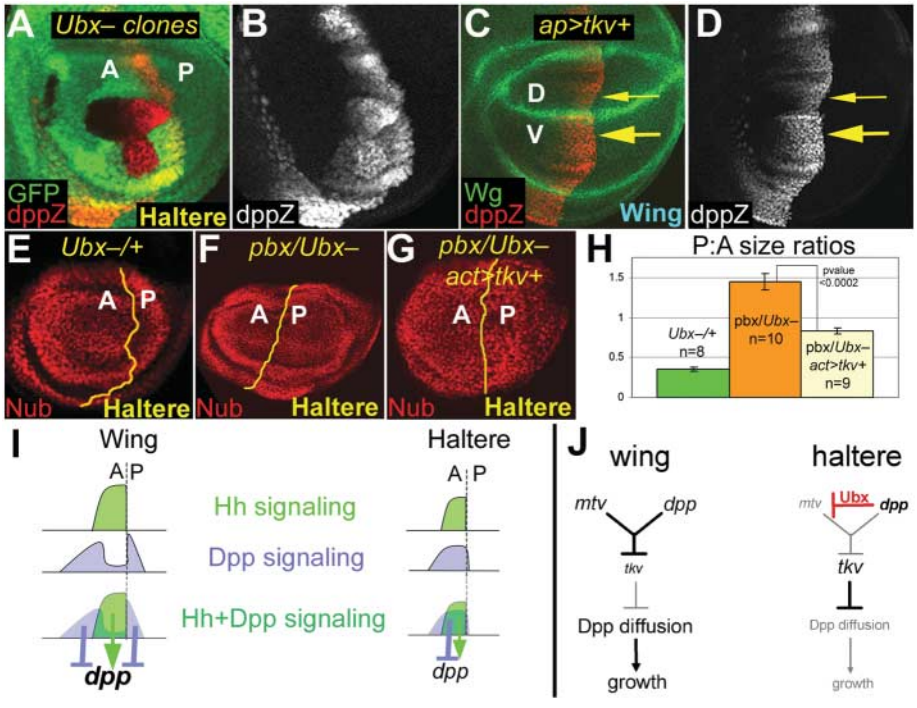


Fig. 5. Contributions of *dpp* transcription and mobility to growth. (**A** and **B**) *dpp-lacZ* is up-regulated in *Ubx* mutant haltere tissue (marked by loss of GFP) within the AP organizer. (**C** and **D**) *UAS-tkv*$^{+}$ expression in dorsal cells using *ap-Gal4* results in *dpp-lacZ* down-regulation (thin arrow) compared with ventral cells (thick arrow). (**E** to **G**) Examples of $+/Ubx^{-}$ (E), pbx/Ubx^{-} (F), and pbx/Ubx^{-}; *actGal4>tkv*$^{+}$ (G) haltere discs stained for Nubbin (Nub), a marker of the appendage, and a marker of the AP compartment boundary (yellow line). (**H**) P:A ratios of the Nub domains of $+/Ubx^{-}$, pbx/Ubx^{-}, and pbx/Ubx^{-}; *actGal4>tkv*$^{+}$ haltere discs. Error bars indicate SEM. (**I**) Summary of the consequences of different spatial relationships of Dpp and Hh signaling for *dpp* transcription in the wing and haltere. (**J**) Summary of how Dpp represses *mtv* in the presence of *Ubx* to control *tkv* expression, Dpp mobility, and growth in the haltere compared with the wing.

haltere levels (fig. S4), and ectopic *mtv* in the haltere did not repress *tkv-lacZ* expression to the extent seen in the medial wing (fig. S4).

Control of the relative position of Dpp and Hh signaling by *tkv* regulation. Because of high Tkv production in the wild-type haltere disc, peak Dpp signal transduction occurs in the AP organizer, the same cells that transduce the Hh signal. Thus, in the haltere, the activity

profiles for these two signal transduction pathways coincide with each other (Fig. 1, F to H, and 5I). In contrast, low *tkv* expression in the wing AP organizer results in two peaks of Dpp signaling that are on either side of Hh-transducing cells. This difference will have important consequences for the expression of genes that are targets of both pathways. For example, *dpp* is activated by Hh and repressed by Dpp signaling (*19*–*22*, *33*). In the haltere, these two conflicting inputs occur in the same cells, possibly contributing to reduced *dpp* expression compared with the wing. Ubx^- clones cell-autonomously up-regulated *dpp-lacZ* in the haltere (Fig. 5, A and B). To test whether *Ubx* lowers *dpp* transcription in part by aligning Dpp and Hh signaling, we expressed uniform *UAS-*tkv^+ in the dorsal half of the wing disc. As a result, in this region of the wing disc both signals peaked in the same cells, and *dpp-lacZ* expression was reduced compared with the ventral half of these wing discs (Fig. 5, C and D, and fig. S5). Conversely, expressing *tkvRNAi* in dorsal haltere cells increased *dpp-lacZ* expression (fig. S5). Thus, *Ubx* reduces *dpp* transcription in part by changing where peak Dpp signaling occurs in the disc (Fig. 5I). We note that *Ubx* is likely to reduce *dpp* expression in additional ways, because increasing *tkv* expression does not lower *dpp-lacZ* expression to that observed in wild-type haltere. Nevertheless, varying the relative spatial relationships between signal transduction pathways is a potentially powerful mechanism for modifying the outputs from commonly used pathways. We suggest that selector genes may work through molecules that control ligand distribution to vary the spatial relationships between these and other signal transduction pathways in diverse contexts during development.

Dpp mobility versus *dpp* transcription. The finding that increased *tkv* expression results in decreased *dpp* transcription reveals an unexpected link between Dpp mobility and Dpp production. Because of this link, the above experiments do not discriminate between growth effects due to differences in Dpp mobility per se as opposed to secondary consequences on Dpp production. To distinguish between these scenarios, we made use of a compartment-specific *Ubx* regulatory allele, *posterior bithorax* (*pbx*), that lacks detectable Ubx in the P compartment when paired with a *Ubx* null allele but still has normal *Ubx* expression in the A compartment (fig. S6) (*8*). Consequently, in *pbx*/Ubx^- haltere discs, the P compartment increased in size such that the P:A size ratio was 1.45 (Fig. 5, E and F); the P:A ratio of +/Ubx^- haltere discs was ~0.35 (*13*). We suggest that Dpp more readily diffuses into and through the P compartments of *pbx*/Ubx^- discs because of the wing-like expression pattern of *tkv* and that this wing-like diffusion results in its robust growth.

To test whether differences in Tkv-regulated Dpp diffusion affect tissue growth independently of an effect on Dpp production, we examined the consequences of expressing *UAS-*tkv^+ uniformly in *pbx*/Ubx^- haltere discs. If Tkv's effect on growth is mediated only by lowering Dpp production, both compartments should be reduced in size and thus maintain the same size ratio. However, if reducing Dpp mobility directly affects growth, the P compartment should be reduced in size more than the A compartment, which, in *pbx*/Ubx^- discs, already has high *tkv* expression. We found that expressing uniform tkv^+ in *pbx*/Ubx^- discs decreased the size of the P compartment more than the A compartment, resulting in a P:A ratio of 0.83 (Fig. 5, E to H). Because uniform tkv^+ returned the P:A ratio back to the wild-type ratio by ~56% (from 1.45 to 0.83, whereas +/Ubx^- discs have a P:A ratio of ~0.35), these results suggest that this single variable is sufficient to provide ~50% rescue of the size of an otherwise *Ubx* mutant P compartment.

Discussion. We have investigated the mechanism underlying a classic yet poorly understood phenomenon in biology: how size variations are genetically programmed in animal development. Many experiments show that organ size is not governed by counting cell divisions but instead depends on disc-intrinsic yet cell-nonautonomous mechanisms, possibly relying on morphogen signaling (*34*). Our results support this idea by showing that alterations in a morphogen gradient contribute to size differences between appendages. In the example investigated here, *Ubx* limits the size of the haltere by reducing both Dpp production and Dpp mobility. Moreover, both of these effects are due, in part, to higher *tkv* expression in the medial haltere (Fig. 5, I and J). In many morphogen systems, the receptors themselves have been shown to control the distribution of the ligand and, consequently, pathway activation (*30*, *35*–*37*). We show that a selector gene exploits this phenomenon to modify organ size.

Although the mechanism by which Dpp controls proliferation is not fully understood, recent results argue that, in the medial wing disc, cells may compare the amount of Dpp transduction with their neighbors, whereas lateral cells proliferate in response to absolute Dpp levels (*17*). Our results suggest several ways in which the altered Dpp gradient in the haltere could limit its growth. First, proliferation of lateral haltere cells may be limited because they perceive less Dpp. Second, the narrower Dpp gradient results in fewer cells exposed to the gradient in the medial haltere. Another notable difference is that, because there are two peaks of Dpp signaling in the wing but only one in the haltere, the wing has four distinct slopes whereas the haltere has only two. The less complex Dpp activity landscape of the haltere may also contribute to its reduced growth.

On the basis of these results, we suggest that altering the shape and intensity of morphogen gradients may be a general mechanism by which selector genes affect tissue sizes in animal development. Consistent with this view, *wingless* (*wg*), another long-range morphogen in the wing, is partially repressed in the haltere (*38*). Intriguingly, some of the size and shape differences in the beaks of Darwin's finches are controlled by alterations in the production of the Dpp ortholog BMP4 (*39*). Our results suggest that differences in the diffusion of this ligand may also contribute to the range of beak morphologies that have evolved in these species.

References and Notes

1. P. R. Grant, B. R. Grant, *Curr. Biol.* **15**, R614 (2005).
2. J. Zakany, D. Duboule, *Cell Tissue Res.* **296**, 19 (1999).
3. D. M. Wellik, M. R. Capecchi, *Science* **301**, 363 (2003).
4. R. Krumlauf, *Cell* **78**, 191 (1994).
5. A. C. Burke, C. E. Nelson, B. A. Morgan, C. Tabin, *Development* **121**, 333 (1995).
6. M. Kessel, P. Gruss, *Cell* **67**, 89 (1991).
7. S. B. Carroll, S. D. Weatherbee, J. A. Langeland, *Nature* **375**, 58 (1995).
8. E. B. Lewis, *Nature* **276**, 565 (1978).
9. P. A. Beachy, S. L. Helfand, D. S. Hogness, *Nature* **313**, 545 (1985).
10. S. M. Cohen, in *The Development of* Drosophila melanogaster, M. Bate, A. Martinez Arias, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993), vol. II, pp. 747–842.
11. P. Martin, *J. Exp. Zool.* **222**, 97 (1982).
12. F. Roch, M. Akam, *Development* **127**, 97 (2000).
13. See Materials and Methods on *Science* Online.
14. G. Morata, A. Garcia-Bellido, W. Roux, *Arch. Dev. Biol.* **179**, 125 (1976).
15. T. Lecuit *et al.*, *Nature* **381**, 387 (1996).
16. D. Nellen, R. Burke, G. Struhl, K. Basler, *Cell* **85**, 357 (1996).
17. D. Rogulja, K. D. Irvine, *Cell* **123**, 449 (2005).
18. P. A. Lawrence, G. Struhl, *Cell* **85**, 951 (1996).
19. K. Basler, G. Struhl, *Nature* **368**, 208 (1994).
20. I. Guillen *et al.*, *Development* **121**, 3447 (1995).
21. T. Tabata, C. Schwartz, E. Gustavson, Z. Ali, T. B. Kornberg, *Development* **121**, 3359 (1995).
22. M. Zecca, K. Basler, G. Struhl, *Development* **121**, 2265 (1995).
23. R. Burke, K. Basler, *Development* **122**, 2261 (1996).
24. J. Capdevila, I. Guerrero, *EMBO J.* **13**, 4459 (1994).
25. J. J. Sekelsky, S. J. Newfeld, L. A. Raftery, E. H. Chartoff, W. M. Gelbart, *Genetics* **139**, 1347 (1995).
26. H. Tanimoto, S. Itoh, P. ten Dijke, T. Tabata, *Mol. Cell* **5**, 59 (2000).
27. E. V. Entchev, A. Schwabedissen, M. Gonzalez-Gaitan, *Cell* **103**, 981 (2000).
28. A. A. Teleman, S. M. Cohen, *Cell* **103**, 971 (2000).
29. T. Y. Belenkaya *et al.*, *Cell* **119**, 231 (2004).
30. T. Lecuit, S. M. Cohen, *Development* **125**, 4901 (1998).
31. T. J. Brummel *et al.*, *Cell* **78**, 251 (1994).
32. Y. Funakoshi, M. Minami, T. Tabata, *Development* **128**, 67 (2001).
33. T. E. Haerry, O. Khalsa, M. B. O'Connor, K. A. Wharton, *Development* **125**, 3977 (1998).
34. S. J. Day, P. A. Lawrence, *Development* **127**, 2977 (2000).
35. K. M. Cadigan, M. P. Fish, E. J. Rulifson, R. Nusse, *Cell* **93**, 767 (1998).
36. J. Casanova, G. Struhl, *Nature* **362**, 152 (1993).
37. Y. Chen, G. Struhl, *Cell* **87**, 553 (1996).
38. S. D. Weatherbee, G. Halder, J. Kim, A. Hudson, S. Carroll, *Genes Dev.* **12**, 1474 (1998).

39. A. Abzhanov, M. Protas, B. R. Grant, P. R. Grant, C. J. Tabin, *Science* **305**, 1462 (2004).
40. Y. S. Lee, R. W. Carthew, *Methods* **30**, 322 (2003).
41. We thank S. Cohen, M. Gonzalez-Gaitan, T. Jessell, L. Johnston, E. Laufer, B. McCabe, M. O'Connor, G. Struhl, T. Tabata, the Bloomington Stock Center, and the Developmental Studies Hybridoma Bank for antibodies and fly stocks and R. Axel, O. Hobert, T. Jessell, L. Johnston, D. Rogulja, and G. Struhl for comments on the manuscript. This work was supported by a grant from the NIH to R.S.M. and an NIH training grant (M.A.C.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/1128650/DC1
Materials and Methods
SOM Text
Figs. S1 to S6

12 April 2006; accepted 25 May 2006
Published online 1 June 2006;
10.1126/science.1128650
Include this information when citing this paper.

Hierarchical Action and Inhibition of Plant Dicer-Like Proteins in Antiviral Defense

Angélique Deleris,[1] Javier Gallego-Bartolome,[1] Jinsong Bao,[2] Kristin D. Kasschau,[2] James C. Carrington,[2] Olivier Voinnet[1]*

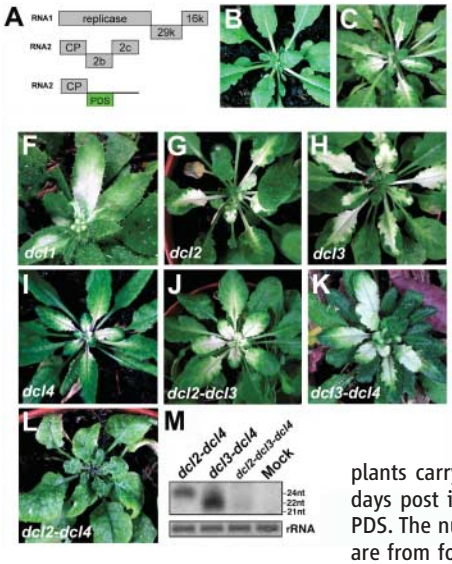
The mechanisms underlying induction and suppression of RNA silencing in the ongoing plant-virus arms race are poorly understood. We show here that virus-derived small RNAs produced by *Arabidopsis* Dicer-like 4 (DCL4) program an effector complex conferring antiviral immunity. Inhibition of DCL4 by a viral-encoded suppressor revealed the subordinate antiviral activity of DCL2. Accordingly, inactivating both DCL2 and DCL4 was necessary and sufficient to restore systemic infection of a suppressor-deficient virus. The effects of DCL2 were overcome by increasing viral dosage in inoculated leaves, but this could not surmount additional, non–cell autonomous effects of DCL4 specifically preventing viral unloading from the vasculature. These findings define a molecular framework for studying antiviral silencing and defense in plants.

In RNA silencing, ribonuclease (RNase) III–like enzymes in the Dicer family produce short interfering (si)RNA and micro (mi)RNA from RNA with double-stranded (ds) features (*1*). These molecules guide RNA-induced silencing complexes (RISCs) to suppress gene expression at the transcriptional, RNA-stability, and translational levels (*2*). *Arabidopsis thaliana* has four specialized Dicer-like (DCL) proteins. DCL1 processes fold-back precursors to release miRNAs (*3*). DCL3 produces 24–nucleotide (nt)–long, DNA repeat–associated siRNAs guiding heterochromatin formation (*4*). DCL4 generates 21-nt-long siRNAs that mediate posttranscriptional silencing of some endogenous genes [trans-acting (ta)–siRNAs; (*5*, *6*)] and of transgenes mediating RNA interference (*7*). DCL2 synthesizes stress-related natural-antisense-transcript (nat)–siRNAs (*8*), siRNAs derived from at least one virus (*4*), and, in *dcl4* mutant plants, it alternately processed ~22-nt siRNAs from ta-siRNA precursors (*5*, *6*).

The observations that virus-derived siRNAs accumulate in plant and insect infected tissues and that many viruses encode suppressor proteins targeting DCL, RISC, or small RNA activities strongly suggest that RNA silencing has antiviral roles (*9*–*11*). In plants, one or more of the six RNA-dependent RNA-polymerase (RDR) paralogs, including *Arabidopsis* RDR6 and RDR1, may strengthen primary silencing responses by producing dsRNA from viral templates (*12*) and by amplifying mobile silencing signals conditioning antiviral immunity in non-infected tissues (*7*, *13*). Nevertheless, the genetic bases of silencing induction and suppression by plant viruses remain unclear. Even the existence of an antiviral RISC ("slicer") is arguable because DCL-mediated processing of virus-derived dsRNA could be, in principle, sufficient to dampen infections. It remains also uncertain how, when, and where antiviral silencing and its suppression impact susceptibility and defense in whole plants. This study addresses these issues using *Arabidopsis* silencing mutants and three distinct RNA viruses.

DCL4- and DCL2-dependent siRNAs recruit an antiviral RISC. *Arabidopsis* plants were inoculated with modified *Tobacco rattle virus* (TRV-PDS) (Fig. 1A) containing a fragment of the *Arabidopsis* phytoene desaturase (*PDS*) gene in place of the RNA2-encoded 2b and 2c sequences. Like TRV-infected tissues (Fig. 1B), TRV-PDS–infected tissues are free of disease symptoms, because of a strong silencing response that dramatically reduces viral titers (*14*), and exhibit extensive photobleaching due to virus-induced gene silencing (VIGS) of *PDS* (Fig. 1C) (*7*).

TRV-PDS–specific siRNAs accumulated as discrete 21-nt and 24-nt species in wild-type (WT) *Arabidopsis* (Fig. 1D), a pattern unchanged in *rdr1*, *rdr2*, *rdr6* [supporting online material (SOM), fig. S1], and *dcl2* mutants (Fig. 1D). However, the 24-nt and 21-nt siRNAs were undetectable in *dcl3* and *dcl4* mutants, respectively. Loss of 21-nt siRNAs coincided with appearance of 22-nt siRNAs in *dcl4* mutants (Fig. 1D). Identical siRNA patterns were detected with an RNA2(TRV)-specific probe, whereas probes specific for cellular *PDS* sequences absent in TRV-PDS



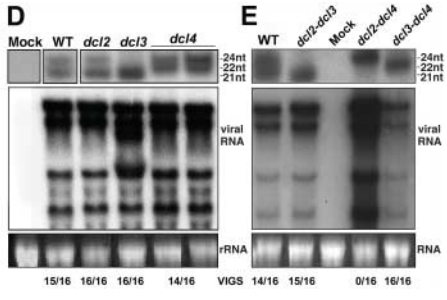


Fig. 1. (**A**) Genome organization of TRV and its TRV-PDS derivative. (**B** and **C**) Asymptomatic infection (B) and extensive photobleaching (C) caused by TRV and TRV-PDS, respectively. (**D** and **E**) Analysis of (top) low- and (bottom) high-molecular-weight RNAs from TRV-PDS–infected plants carrying single (D) and double (E) *dcl* mutations [14 days post inoculation (dpi)]. The probe was specific for viral PDS. The numbers of infected plants showing photobleaching are from four independent experiments involving four plants each. (**F** to **L**) Disease symptoms and VIGS in *dcl* mutants (14 dpi). (**M**) TRV-PDS siRNA analysis in *dcl2-dcl3-dcl4* triple mutants (14 dpi). rRNA shown by ethidium bromide staining.

[1]Institut de Biologie Moléculaire des Plantes, CNRS Unité Propre de Recherche (UPR) 2357, 12, rue du Général Zimmer, 67084 Strasbourg Cedex, France. [2]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 97331, USA.

*To whom correspondence should be addressed. E-mail: olivier.voinnet@ibmp-ulp.u-strasbg.fr

yielded no signal, which indicated that all siRNA species detected were of viral origin (fig. S1). Viral RNA accumulation was not altered in any of the single *dcl* or *rdr* mutants, nor was the extent and/or consistency of VIGS compared with WT-infected plants (Fig. 1, D and F to I) (fig. S1), which suggested redundancy among those factors in mediating antiviral silencing.

dcl combination mutants were then infected. Like *dcl3*, *dcl2-dcl3* double mutants accumulated only 21-nt siRNAs and had unaltered TRV-PDS levels and VIGS phenotype (Fig. 1, E and J). Similar infection and VIGS phenotypes were detected in *dcl3-dcl4* double mutants, although they accumulated exclusively 22-nt siRNAs (Fig. 1, E and K). By contrast, VIGS was abolished in *dcl2-dcl4* plants accumulating only 24-nt siRNAs. This coincided with higher virus RNA levels and stronger disease symptoms (Fig. 1, E and L), also observed in *dcl2-dcl3-dcl4* triple mutants, in which low-abundance, DCL1-dependent siRNAs were detected (Fig. 1M; fig. S1).

Although the DCL3-dependent, 24-nt siRNA accumulated to the same high levels as 21-nt and 22-nt siRNAs (Fig. 1, D and E), it was neither necessary (Fig. 1D) nor sufficient (Fig. 1E) to mediate VIGS and defense against TRV. Thus, the DCL3-dependent dicing reaction alone could not limit virus infection. The additional requirement for an siRNA-loaded RISC was evidenced by the fact that loss of slicer activity (i.e., *PDS* VIGS), appearance of disease symptoms, and high virus accumulation were inherently correlated in *dcl2-dcl4*. We conclude that the respective 21-nt and 22-nt siRNA products of DCL4 and DCL2 (Fig. 1, D and E) guide an antiviral RISC to promote defense against TRV. DCL2 likely acts as a DCL4 substitute because its activity was contingent on DCL4 loss of function.

TCV-encoded P38 suppresses DCL4. To assess the generality of these findings and the significance of a DCL4 substitute, we analyzed infection by *Turnip crinkle virus* (TCV, Fig. 2A). Unlike TRV, TCV causes disease symptoms in WT *Arabidopsis* (fig. S2) and encodes a strong silencing suppressor, the P38 capsid protein (*15*). In WT *Arabidopsis*, TCV-derived siRNAs accumulated as a single, 22-nt species. This profile was not changed in *dcl3*, *dcl4*, or *dcl3-dcl4* mutants (Fig. 2A) and not in *rdr1*, *rdr2*, or *rdr6* backgrounds (fig. S2). By contrast, 22-nt siRNA levels were strongly reduced in *dcl2* mutants, as has been reported (*4*), and in *dcl2-dcl3* and *dcl2-dcl4* double mutants (Fig. 2A). A time-course analysis revealed that all single, double, and triple mutants with *dcl2* mutations accumulated statistically higher TCV RNA levels than did WT plants or mutants without *dcl2* (Fig. 2B). However, these effects were modest (up to twofold), as also revealed by P38 immunoblots and symptom evaluation (Fig. 2A and fig. S2), and were disproportionate compared with the near-loss of 22-nt siRNAs in *dcl2* and its derivatives.

Previous studies suggested that DCL2 functions redundantly with at least one other DCL (*4*), and we envisaged that P38 effectively masked the effects of this second TCV-antagonizing activity. To test this hypothesis, we analyzed P38-expressing *Arabidopsis* that also expressed a second inverted-repeat (IR) transgene producing siRNAs against the chalcone synthase (*CHS*) mRNA (Fig. 2C). Transgenic P38 restored *CHS* accumulation, significantly reduced 21-nt *CHS* siRNA levels, and triggered accumulation of less abundant, 22-nt siRNAs (Fig. 2C, left panel). This resembled the TRV-PDS siRNA patterns in DCL4-deficient mutants (Fig. 1, D and E), which suggested that P38 suppresses DCL4. Accordingly, endogenous DCL4-dependent ta-siRNAs were specifically lost in P38 plants (Fig. 2C, left). TCV infection recapitulated all these DCL4-suppressing effects (Fig. 2C, right). These data suggested that both DCL4 and DCL2 mediate TCV silencing, with DCL2 providing redundant siRNA-processing functions when DCL4 is suppressed by P38.

DCL4 and DCL2 redundantly silence P38-deficient TCV. To fully appreciate DCL4 effects on TCV silencing, we used a recombinant virus (in vitro transcript inoculum of 500 ng/leaf) in which the *green fluorescent protein* (*gfp*) reporter-gene replaced the P38 sequence [TCV-GFPΔP38, Fig. 3A and (*16*)]. TCV-GFPΔP38 siRNAs accumulated as a 21-nt species in inoculated leaves of WT plants and single *dcl* mutants, except in *dcl4*, in which siRNAs were 22 nt long (Fig. 3A). Viral RNA levels were low in *dcl2* and *dcl3* mutants, but higher in *dcl4* and *dcl3-dcl4* plants (Fig. 3B), which suggested a greater antiviral contribution of DCL4 alone in TCV-GFPΔP38 than in TRV-PDS infections (Fig. 1D). TCV-GFPΔP38 levels were extremely high in *dcl2-dcl4* (Fig. 3B), consistent with results of *dcl2-dcl4* and *dcl2-dcl3-dcl4* mutants infected with wild-type TCV (Fig. 2B). Although 24-nt siRNAs were undetectable in WT-infected plants, they accumulated in *dcl2-dcl4* plants (Fig. 3C), as in TRV-PDS–infected plants (Fig. 1E). Also as with TRV-PDS, very low siRNA levels were detected in *dcl2-dcl3-dcl4*, likely reflecting inefficient DCL1 activities (Fig. 3C). We conclude that a similar DCL consortium affected siRNA production and virus levels in plants infected with TRV-PDS or TCV-GFPΔP38, and that DCL4 and DCL2 redundantly mediate defense against P38-deficient TCV.

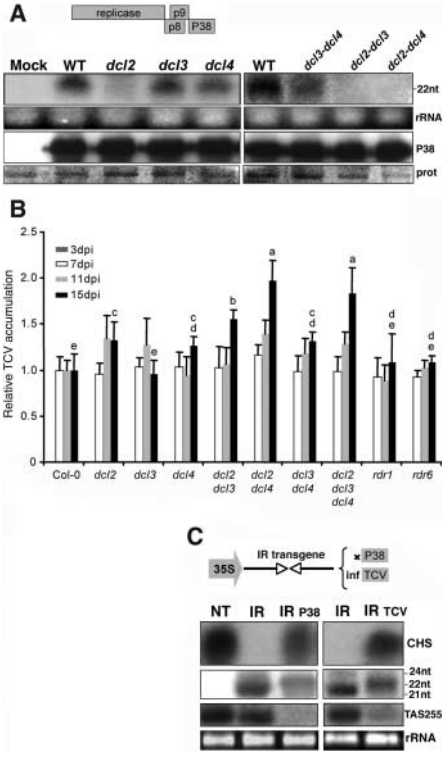


Fig. 2. (**A**) TCV genome organization; TCV siRNA (top) and coat-protein (P38) accumulations (bottom) in single or double *dcl* mutants (14 dpi). Prot, total protein staining. (**B**) Time-course analysis of TCV genomic RNA accumulation. Variance analysis of the data from 15 dpi produced five statistical groups (a to e). (**C**) Analysis of high (top) and low (second from top) molecular-weight RNAs derived from the inverted-repeat (IR) *CHS* locus in nontransgenic (NT), P38-transgenic (left) or TCV-infected plants (right). (Bottom) Accumulation of TAS255.

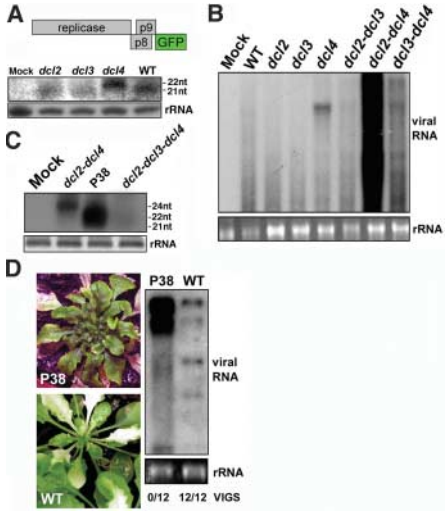


Fig. 3. (**A**) Genome organization of TCV-GFPΔP38 and TCV-GFPΔP38 siRNA analysis in *dcl* mutants (in vitro transcripts; 500 ng/leaf; 7 dpi). (**B**) TCV-GFPΔP38 RNA accumulation in single and double *dcl* mutants. The probe was GFP-specific. (**C**) TCV-GFPΔP38 siRNA analysis in double and triple *dcl* mutants and P38-expressing plants. (**D**) TRV-PDS–induced photobleaching occurs in WT, but not P38-expressing plants (14 dpi). (Right) TRV-PDS RNA accumulation. The numbers of infected plants showing photobleaching are from four independent experiments involving three plants each.

The loss of DCL2 had only limited impact on TCV susceptibility (Fig. 2, A and B). Likewise, *CHS* silencing was released in P38-expressing and TCV-infected plants, despite accumulation of 22-nt siRNAs (Fig. 2C). This suggested that 22-nt siRNAs had suboptimal antiviral activity in the presence of P38. Accordingly, transgenic P38 prevented VIGS, promoted strong disease symptoms and high accumulation of TRV-PDS (Fig. 3D; fig. S2), effects that specifically required the combined inactivation of DCL2 and DCL4 in nontransgenic plants (Fig. 1, E and L). We conclude that besides primary DCL4-antagonizing activities, P38 also likely suppresses the action of DCL2-dependent siRNAs, consistent with major antiviral roles for both enzymes. Further demonstrating the key antiviral functions of DCL4 and DCL2, Cucumber mosaic virus (CMV) levels were three to four times those in *dcl2-dcl4* plants compared with WT or single *dcl2* or *dcl4* mutants. Unlike TCV and TRV, however, CMV silencing was dependent on RDR6 (*14*), loss of which resulted in CMV levels that were indistinguishable from those in *dcl2-dcl4* double mutants (fig. S3).

DCL2-DCL4–dependent block to systemic virus movement. We exploited the GFP tag to follow TCV-GFPΔP38 infection in WT and *dcl* mutant plants, initially using a moderate in vitro transcript inoculum (500 ng/leaf). TCV-GFPΔP38 moved from cell to cell to form scattered primary lesions in WT-inoculated leaves (Fig. 4A, inlay) but consistently failed to spread systemically and produced no disease symptom (Fig. 4, B and C) (*16*). Those defects were rescued in P38 transgenic plants (Fig. 4, D to F), which exhibited large, confluent primary lesions in inoculated leaves (Fig. 4D). The aborted systemic spread of TCV-GFPΔP38 could have resulted from its failure to counteract the DCL4-DCL2 antiviral effects, owing to the lack of P38. However, because P38 is also the viral capsid, it could have simply resulted from an inability to form virions, a prerequisite to systemic infection of many plant viruses (*17*).

To address these issues, we used the *Arabidopsis* amplicon line (AMP, in SOM Text) in which replication of a GFP-tagged RNA virus triggers silencing of viral transcripts and low GFP accumulation (Fig. 4, G and H). Transgenic P38 (line AMP-P38) and TCV infection suppressed amplicon silencing [Fig. 4, G (middle and bottom), and H] and inhibited accumulation of endogenous ta-siRNA255 (Fig. 4H). A genetic screen for loss of the AMP-P38 phenotype identified two mutants, m1 and m2, that exhibited reduced amplicon RNA accumulation and GFP expression and also contained WT ta-siRNA255 levels (Fig. 4H). Linkage analysis, DNA sequencing and immunoblot assays revealed that m1 and m2 were stable, point-mutation alleles of P38 (Fig. 4, H and I). The data presented are for m1; similar results were obtained with m2.

Despite extensive cell-to-cell movement in inoculated leaves to levels resembling those of virion-proficient P38 plants (Fig. 4D), long-distance spread of TCV-GFPΔP38 was prevented in plants having m1 or m2 mutations in which the AMP locus had been segregated away (Fig. 4J, top). However, both P38 alleles formed virions, because TCV-GFPΔP38 was transmitted from infected plants with m1 or m2 mutations to healthy, P38-expressing plants through sap inoculation, a procedure whereby viral RNAs that are not encapsidated are rapidly degraded (Fig. 4J, bottom). Thus, m1 or m2 genetically uncoupled the virion-packaging and suppression of silencing functions of P38. Therefore, the lack of systemic TCV-GFPΔP38 P38 movement in WT plants was not caused by its inability to form virions, but rather, by its likely failure to counteract the DCL4-DCL2 antiviral effects. To test this idea, a moderate TCV-GFPΔP38 inoculum (500 ng/leaf) was applied to *dcl2*, *dcl4*, and *dcl2-dcl4* mutants. Whereas the infection phenotype of *dcl2* mutants was indistinguishable from that of WT plants, *dcl4* mutants exhibited large primary lesions (Fig. 4K, left) but failed to support systemic movement (Fig. 4K, right). By contrast, there were confluent primary lesions, systemic movement, and disease symptoms in *dcl2-dcl4* plants (Fig. 4L). We conclude that, with a moderate inoculum, the combined loss of DCL4 and DCL2 functions was necessary and sufficient to recapitulate the P38-mediated rescue of P38-deficient virus infection.

DCL4 imposes a barrier to vascular exit of TCV-GFPΔP38. The continuum of infection phenotypes in WT, *dcl4*, and *dcl2-dcl4* plants suggested that the primary effect of DCL2 and DCL4 was to restrict the virus to inoculated leaves (Fig. 4, A, K, and L). A fivefold increase in TCV-GFPΔP38 inoculum (5×, 2.5 μg/leaf) restored systemic movement in *dcl4*-, *dcl2-dcl4*–, and *dcl2-dcl3-dcl4*–inoculated plants, but in none of the other single or combination mutants (Fig. 5, A to C). Moreover, primary infection foci in *dcl4*-inoculated leaves were confluent, as in P38-expressing plants (Fig. 5C). By contrast, and despite formation of macroscopic primary foci, the same (2.5 μg/leaf, Fig. 4D) or much higher-dose (25×, 12.5 μg/leaf) inocula failed to promote systemic movement in WT plants, even though the virus consistently entered the petiole vasculature (Fig. 5, D and E, arrows). Therefore, increasing the inoculum strength could bypass the antiviral effects of DCL2 but not those of DCL4.

To gain insight into additional antiviral roles of DCL4, TCV-GFPΔP38–inoculated leaves were examined in greater detail. GFP was evenly distributed in petioles of *dcl4* plants (Fig. 5C, arrow), whereas in similar WT tissues it appeared as thin stripes (Fig. 5, D and E, arrows), identified as vascular bundles (Fig. 5, F to I). In contrast, TCV-GFPΔP38 could readily exit the vasculature of *dcl4* petioles, invading parenchyma and epidermal cells (Fig. 5, J and K). Thus, although the virus penetrated the phloem of WT plants, its unloading was

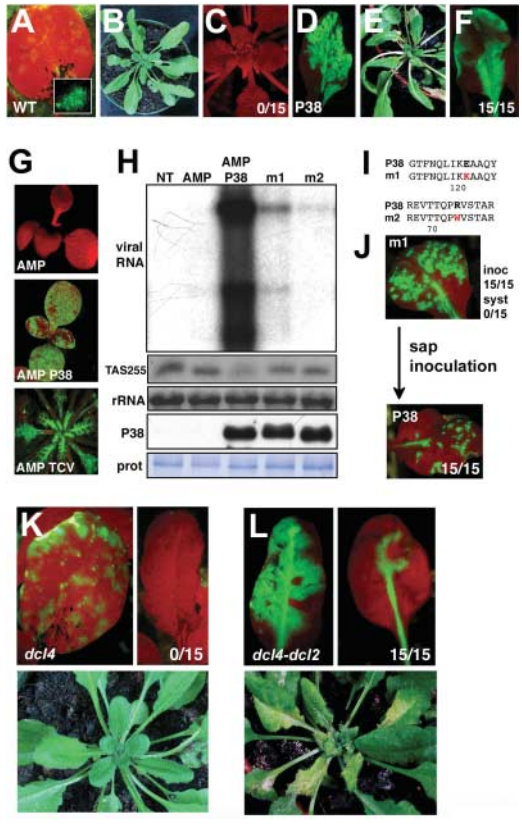


Fig. 4. (**A**) WT *Arabidopsis* leaf challenged with moderate TCV-GFPΔP38 inoculum (in vitro transcripts, 500 ng/leaf; 7 dpi). Chlorophyll fluoresces red under UV. (Inlay) Microscopic primary lesion. (**B** and **C**) Lack of symptoms (B) and of systemic virus movement (C) in the plant in (A). The numbers of infected plants showing systemic viral GFP are from three independent experiments involving five plants each. (**D** to **F**) Same as (A to C), but in P38-expressing plants. (**G**) GFP silencing in the amplicon (AMP, top) is released by transgenic P38 (AMP-P38) and TCV infection (middle and bottom). (**H**) Analysis of high- and low-molecular-weight RNAs from nontransgenic (NT), AMP, and AMP-P38 and from m1 and m2 mutants recovered after AMP-P38 mutagenesis. (Bottom) P38 immunoblot analysis. (**I**) Amino acid substitutions in m1 and m2 alleles of P38. (**J**) Confluent primary lesions and lack of systemic viral movement in m1 after high-dose inoculation (2.5 μg/leaf; top). TCV-GFPΔP38 is 100% transmitted from m1 to P38-expressing plants through sap extracts (bottom). (**K**) Same as (A to C) but in *dcl4* mutants. (**L**) Same as (A to C) but in *dcl2-dcl4* (14 dpi).

specifically prevented by a DCL4-dependent mechanism. This mechanism was likely not cell autonomous, because it was also manifested in petioles of high-dose inoculated WT leaves (12.5 μg inoculum per leaf) in which virus accumulation was likely as elevated as in *dcl4* tissues (Fig. 5, E compared with C).

We recently identified DCL4 as an essential component of non–cell autonomous RNA silencing in a genetic screen using the *SUC-SUL* system, whereby a vascular-specific IR transgene triggers *SUL* silencing movement, resulting in vein-centered chlorosis [(*7*), Fig. 5L]. Of the two *SUC-SUL*–derived siRNAs species (21-nt and 24-nt; Fig. 5L), only DCL4-dependent 21-nt siRNAs were found required for cell-to-cell signaling (*7*). We reasoned that the vein-restriction of TCV-GFPΔP38 could have resulted from its failure to suppress the effects of an antiviral, DCL4-dependent silencing signal (Fig. 5Q). If so, P38 and TCV were expected to inhibit vascular exit of silencing signals. Transgenic P38 and TCV infection indeed alleviated silencing movement in *SUC-SUL* plants (Fig. 5, M to O). Moreover, the 21-nt siRNA levels were reduced in TCV-infected tissues (Fig. 5P), recapitulating the effects of *dcl4* knockout mutations that specifically prevented *SUL* silencing movement (*7*). Collectively, these results support the ideas that DCL4 exerts some of its antiviral effects by producing a silencing signal that restricts virus exit from vascular bundles, and that vascular production of P38 inhibits this signaling during normal infections (Fig. 5Q).

Conclusions. Although virus-derived dsRNA is accessible to each of the four DCLs under appropriate genetic circumstances, DCL4 and DCL2 exhibit specific, hierarchical antiviral activities. The primary sensor is DCL4, which produces 21-nt siRNAs that guide a virus-antagonizing RISC. Secondarily, DCL2 forms 22-nt siRNAs with antiviral activities, but these are manifested prominently when DCL4 is genetically removed or suppressed. Hypersusceptibility was only evident when both enzymes were inactivated, which revealed their combined action in defense. Despite detection of DCL3 and, to a limited extent, DCL1 activities on virus-derived substrates, these were not associated with antiviral defense. The hierarchical DCL access to virus-derived dsRNAs is similar to that detected with ta-siRNA precursors: Ta-siRNA biogenesis involves the preferential activity of DCL4, but both DCL3 and DCL2 gain access to precursors in *dcl4* mutants (*5*, *6*). Such specificity and hierarchical action possibly reflect contrasted DCL affinities for distinct dsRNA substrates and are likely to apply to viruses impacted by RDR6, such as CMV, and to viruses with DNA genomes (*18*).

Previous studies suggested that P38 suppresses one DCL (*15*), and we show that DCL4 is the primary target. However, DCL4 suppression was rescued by DCL2, although the antiviral activity of 22-nt siRNAs was in turn compromised by P38. Have DCL4 and DCL2 evolved primarily for optimal, redundant processing of various forms of pathogenic dsRNA? The existence of endogenous ta-siRNAs and nat-siRNAs that involve DCL4 and DCL2, respectively, complicates this question (*5*, *6*, *8*). However, exposure to highly diverse viral suppressors that impact DCL functions may explain why the DCL family has proliferated in plants and why DCL4 and DCL2 diverge in sequence much faster than the miRNA-specific DCL1 (SOM Text). Nevertheless, several silencing suppressors directly sequester DCL products, and these may differently influence evolution of silencing components (*19*).

Finally, this study indicates that a full appreciation of plant antiviral silencing will require analysis of viral genomes with disabled silencing suppressor functions. This notably uncovered the specific effects of DCL4 on virus vascular exit. The finding that TCV movement required the suppressor function of P38 independently of its virion-packaging function sheds a new light on systemic movement by offering a molecular explanation as to why the bundle sheath–phloem interface usually acts as a key boundary against long-distance transport of most plant viruses (*17*).

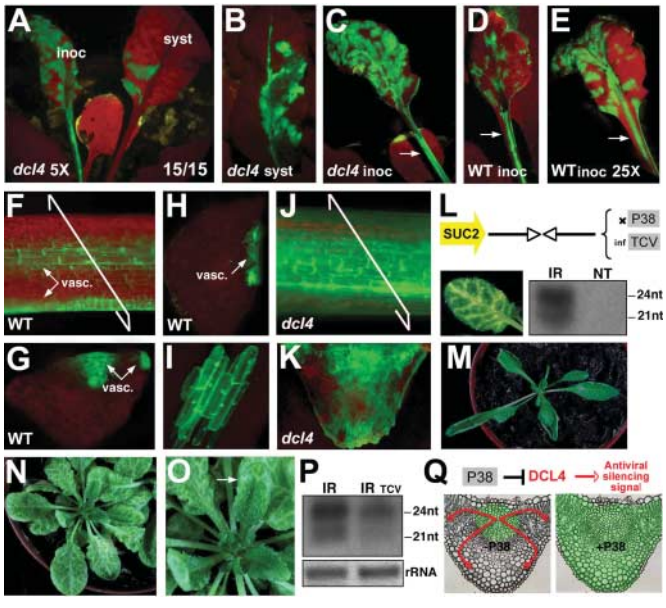


Fig. 5. (**A**) Systemic (syst) TCV-GFPΔP38 movement in *dcl4* plants challenged with high-dose inoculum (5×, 2.5 μg/leaf; inoc). (**B**) TCV-GFPΔP38 unloading in systemic *dcl4* leaves. (**C**) Confluent primary lesions and uniform GFP distribution (arrow) in petioles of *dcl4*-inoculated leaves. (**D**) Same as (C) but in WT plants. (**E**) The TCV-GFPΔP38 inoculum was five times that shown in (A) (25×, 12.5 μg/leaf). (**F**) Magnified view of the petiole in (E). Vasc, vascular bundles. (**G** and **H**) Transverse sections of (F). (**I**) Longitudinal section of (G) showing GFP confined into phloem cells. (**J**) Same as (F) but in *dcl4*-inoculated leaves. (**K**) Transverse section of (J). (**L**) The inverted-repeat (IR) *SUC-SUL* system. SUC2, phloem-specific promoter. (**M**) Transgenic P38 expression inhibits *SUL* silencing movement. (**N**) Loss of silencing movement in young tissues of TCV-infected *SUC-SUL* plants. (**O**) Magnified view of the apex in (N). Arrow, typical sink- (bottom leaf part, containing TCV) to-source (upper leaf part, not containing TCV) transition, showing that suppression of silencing movement is contingent on viral invasion. (**P**) *SUL* siRNA analysis in tissues depicted in (O). (**Q**) Model for dcl4-dependent vascular restriction of TCV-GFPΔP38 and its suppression by P38.

References and Notes

1. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
2. D. Baulcombe, *Nature* **431**, 356 (2004).
3. D. P. Bartel, *Cell* **116**, 281 (2004).
4. Z. Xie *et al.*, *PLoS Biol.* **2**, E104 (2004).
5. V. Gasciolli, A. C. Mallory, D. P. Bartel, H. Vaucheret, *Curr. Biol.* (2005).
6. Z. Xie, E. Allen, A. Wilken, J. C. Carrington, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12984 (2005).
7. P. Dunoyer, C. Himber, O. Voinnet, *Nat. Genet.* **37**, 1356 (2005).
8. O. Borsani, J. Zhu, P. E. Verslues, R. Sunkar, J. K. Zhu, *Cell* **123**, 1279 (2005).
9. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (1999).
10. O. Voinnet, *Nat. Rev. Genet.* **6**, 206 (2005).
11. X. H. Wang *et al.*, *Science* **312**, 452 (2006).
12. P. Mourrain *et al.*, *Cell* **101**, 533 (2000).
13. F. Schwach, F. E. Vaistij, L. Jones, D. C. Baulcombe, *Plant Physiol.* **138**, 1842 (2005).
14. F. Ratcliff, B. D. Harrison, D. C. Baulcombe, *Science* **276**, 1558 (1997).
15. F. Qu, T. Ren, T. J. Morris, *J. Virol.* **77**, 511 (2003).
16. Y. Cohen, A. Gisel, P. C. Zambryski, *Virology* **273**, 258 (2000).
17. H. B. Scholthof, *Trends Plant Sci.* **10**, 376 (2005).
18. R. Akbergenov *et al.*, *Nucleic Acids Res.* **34**, 462 (2006).
19. L. Lakatos *et al.*, *EMBO J.* **25**, 2768 (2006).

20. We thank members of the Voinnet laboratory for discussions and Z. Xie for *dcl* seeds. Funded by CNRS grant to A.D.; NSF grant MCB-0209836, NIH grant AI43288, and U.S. Department of Agriculture grant NRI 2005-35319-15280 to J.C.; and Pao Schloarship (Zhejiang University, China) to J.B. This work is dedicated to the memory of M. and G. Voinnet.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1128214/DC1
Materials and Methods
SOM Text
Figs. S1 to S3
References and Notes

3 April 2006; accepted 18 May 2006
Published online 1 June 2006;
10.1126/science.1128214
Include this information when citing this paper.

The Geometry of Musical Chords

Dmitri Tymoczko

A musical chord can be represented as a point in a geometrical space called an orbifold. Line segments represent mappings from the notes of one chord to those of another. Composers in a wide range of styles have exploited the non-Euclidean geometry of these spaces, typically by using short line segments between structurally similar chords. Such line segments exist only when chords are nearly symmetrical under translation, reflection, or permutation. Paradigmatically consonant and dissonant chords possess different near-symmetries and suggest different musical uses.

Western music lies at the intersection of two seemingly independent disciplines: harmony and counterpoint. Harmony delimits the acceptable chords (simultaneously occurring notes) and chord sequences. Counterpoint (or voice leading) is the technique of connecting the individual notes in a series of chords so as to form simultaneous melodies. Chords are usually connected so that these lines (or voices) move independently (not all in the same direction by the same amount), efficiently (by short distances), and without voice crossings (along nonintersecting paths) (Fig. 1, A to C). These features facilitate musical performance, engage explicit aesthetic norms (*1*, *2*), and enable listeners to distinguish multiple simultaneous melodies (*3*).

How is it that Western music can satisfy harmonic and contrapuntal constraints at once? What determines whether two chords can be connected by efficient voice leading? Musicians have been investigating these questions for almost three centuries. The circle of fifths (fig. S1), first published in 1728 (*4*), depicts efficient voice leadings among the 12 major scales. The Tonnetz (fig. S2), originating with Euler in 1739, represents efficient voice leadings among the 24 major and minor triads (*2*, *5*). Recent work (*5*–*13*) investigates efficient voice leading in a variety of special cases. Despite tantalizing hints (*6*–*10*), however, no theory has articulated general principles explaining when and why efficient voice leading is possible. This report provides such a theory, describing geometrical spaces in which points represent chords and line segments represent voice leadings between their endpoints. These spaces show us precisely how harmony and counterpoint are related.

Human pitch perception is both logarithmic and periodic: Frequencies f and $2f$ are heard to be separated by a single distance (the octave) and to possess the same quality or chroma. To model the logarithmic aspect of pitch perception, I associate a pitch's fundamental frequency f with a real number p according to the equation

$$p = 69 + 12 \log_2(f/440) \qquad (1)$$

The result is a linear space (pitch space) in which octaves have size 12, semitones (the distance between adjacent keys on a piano) have size 1, and middle C is assigned the number 60. Distance in this space reflects physical distance on keyboard instruments, orthographical distance in Western musical notation, and musical distance as measured in psychological experiments (*14*, *15*).

Musically, the chroma of a note is often more important than its octave. It is therefore useful to identify all pitches p and $p + 12$. The result is a circular quotient space (pitch-class space) that mathematicians call $\mathbb{R}/12\mathbb{Z}$ (fig. S3). (For a glossary of terms and symbols, see table S1.) Points in this space (pitch classes) provide numerical alternatives to the familiar letter-names of Western music theory: C = 0, C♯/D♭ = 1, D = 2, D quarter-tone sharp = 2.5, etc. Western music typically uses only a discrete lattice of points in this space. Here I consider the more general, continuous case. This is because the symmetrical chords that influence voice-leading behavior need not lie on the discrete lattice.

The content of a collection of notes is often more important than their order. Chords can therefore be modeled as multisets of either pitches or pitch classes. ("Chord" will henceforth refer to a multiset of pitch classes unless otherwise noted.) The musical term "transposition" is synonymous with the mathematical term "translation" and is represented by addition in pitch or pitch-class space. Transpositionally related chords are the same up to translation; thus, the C major chord, {C, E, G} or {0, 4, 7}, is transpositionally related to the F major chord, {F, A, C} or {5, 9, 0}, because {5, 9, 0} ≡ {0 + 5, 4 + 5, 7 + 5} modulo $12\mathbb{Z}$. The musical term "inversion" is synonymous with the mathematical term "reflection" and corresponds to subtraction from a constant value. Inversionally related chords are the same up to reflection; thus, the C major chord is inversionally related to the C minor chord {C, E♭, G}, or {0, 3, 7}, because {0, 3, 7} ≡ {7 – 7, 7 – 4, 7 – 0} modulo $12\mathbb{Z}$. Musically, transposition and inversion are important because they preserve the character of a chord: Transpositionally related chords sound extremely similar, inversionally related chords fairly so (movie S1).

A voice leading between two multisets $\{x_1, x_2, \ldots, x_m\}$ and $\{y_1, y_2, \ldots, y_n\}$ is a multiset of ordered pairs (x_i, y_j), such that every member of each multiset is in some pair. A trivial voice leading contains only pairs of the form (x, x). The notation $(x_1, x_2, \ldots, x_n) \rightarrow (y_1, y_2, \ldots, y_n)$ identifies the voice leading that associates the corresponding items in each list. Thus, the voice leading (C, C, E, G) → (B, D, F, G) associates C with B, C with D, E with F, and G with G. Music theorists have proposed numerous ways of measuring voice-leading size. Rather than adopting one, I will require only that a measure satisfy a few constraints reflecting widely acknowledged features of Western music (*16*). These constraints make it possible to identify, in polynomial time, a minimal voice leading

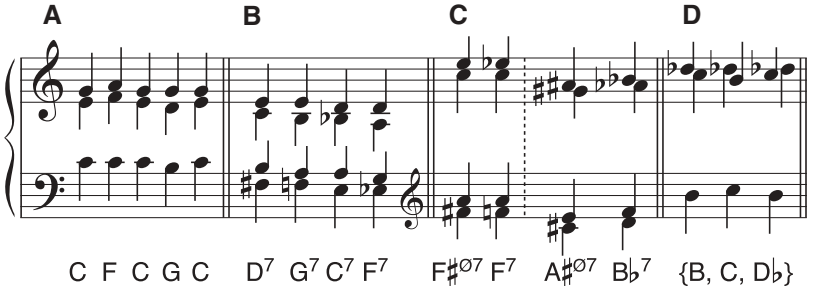


Fig. 1. Efficient voice leading between transpositionally and inversionally related chords. These progressions exploit three near-symmetries: transposition (**A** and **B**), inversion (**C**), and permutation (**D**). Sources: a common 18th-century upper-voice voice-leading pattern (A), a common jazz-piano voice-leading pattern, which omits chord roots and fifths and adds ninths and thirteenths (B), Wagner's *Parsifal* (C), Debussy's *Prélude à l'après-midi d'un faune* (C), and contemporary atonal voice leading in the style of Ligeti and Lutoslawski (D) (soundfile S1). Chord labels refer to unordered sets of pitch classes and do not indicate register.

Department of Music, Princeton University, Princeton, NJ 08544, USA, and Radcliffe Institute for Advanced Study, 34 Concord Avenue, Cambridge, MA 02138, USA. E-mail: dmitri@princeton.edu

(not necessarily bijective) between arbitrary chords (*16*). Every music-theoretical measure of voice-leading size satisfies these constraints.

I now describe the geometry of musical chords. An ordered sequence of n pitches can be represented as a point in $\mathbb{R}^n$ (fig. S4). Directed line segments in this space represent voice leadings. A measure of voice-leading size assigns lengths to these line segments. I will use quotient spaces to model the way listeners abstract from octave and order information. To model an ordered sequence of n pitch classes, form the quotient space $(\mathbb{R}/12\mathbb{Z})^n$, also known as the n-torus $\mathbb{T}^n$. To model unordered n-note chords of pitch classes, identify all points $(x_1, x_2, \ldots x_n)$ and $(x_{\sigma(1)}, x_{\sigma(2)}, \ldots x_{\sigma(n)})$, where σ is any permutation. The result is the global-quotient orbifold $\mathbb{T}^n/\mathcal{S}_n$ (*17*, *18*), the n-torus $\mathbb{T}^n$ modulo the symmetric group $\mathcal{S}_n$. It contains singularities at which the local topology is not that of $\mathbb{R}^n$.

Figure 2 shows the orbifold $\mathbb{T}^2/\mathcal{S}_2$, the space of unordered pairs of pitch classes. It is a Möbius strip, a square whose left edge is given a half twist and identified with its right. The orbifold is singular at its top and bottom edges, which act like mirrors (*18*). Any bijective voice leading between pairs of pitches or pairs of pitch classes can be associated with a path on Fig. 2 (movie S2). Measures of voice-leading size determine these paths' lengths. They are the images of line segments in the parent spaces $\mathbb{T}^n$ and $\mathbb{R}^n$, and are either line segments in the orbifold or "reflected" line segments that bounce off its singular edges. For example, the voice leading (C, D♭) → (D♭, C) reflects off the orbifold's upper mirror boundary (Fig. 2).

Generalizing to higher dimensions is straightforward. To construct the orbifold $\mathbb{T}^n/\mathcal{S}_n$, take an n-dimensional prism whose base is an $(n - 1)$ simplex, twist the base so as to cyclically permute its vertices, and identify it with the opposite face (figs. S5 and S6) (*16*). The boundaries of the orbifold are singular, acting as mirrors and containing chords with duplicate pitch classes. Chords that divide the octave evenly lie at the center of the orbifold and are surrounded by the familiar sonorities of Western tonality. Voice leadings parallel to the height coordinate of the prism act as transpositions. A free computer program written by the author allows readers to explore these spaces (*19*).

In many Western styles, it is desirable to find efficient, independent voice leadings between transpositionally or inversionally related chords. The progressions in Fig. 1 are all of this type (movie S3). A chord can participate in such progressions only if it is nearly symmetrical under transposition, permutation, or inversion (*16*). I conclude by describing these symmetries, explaining how they are embodied in the orbifolds' geometry, and showing how they have been exploited by Western composers.

A chord is transpositionally symmetrical (T-symmetrical) if it either divides the octave into equal parts or is the union of equally sized subsets that do so (*20*). Nearly T-symmetrical chords are close to these T-symmetrical chords. Both types of chord can be linked to at least some of their transpositions by efficient bijective voice leadings. As one moves toward the center of the orbifold, chords become increasingly T-symmetrical and can be linked to their transpositions by increasingly efficient bijective voice leading. The perfectly even chord at the center of the orbifold can be linked to all of its transpositions by the smallest possible bijective voice leading; a related result covers discrete pitch-class spaces (*16*). Efficient voice leadings between perfectly T-symmetrical chords are typically not independent. Thus, composers have reason to prefer near T-symmetry to exact T-symmetry.

It follows that the acoustically consonant chords of traditional Western music can be connected by efficient voice leading. Acoustic consonance is incompletely understood; however, theorists have long agreed that chords approximating the first few consecutive elements of the harmonic series are particularly consonant, at least when played with harmonic tones (*21*). Because elements n to $2n$ of the harmonic series evenly divide an octave in frequency space, they divide the octave nearly evenly in log-frequency space. These chords are therefore clustered near the center of the orbifolds (Table 1) and can typically be linked by efficient, independent

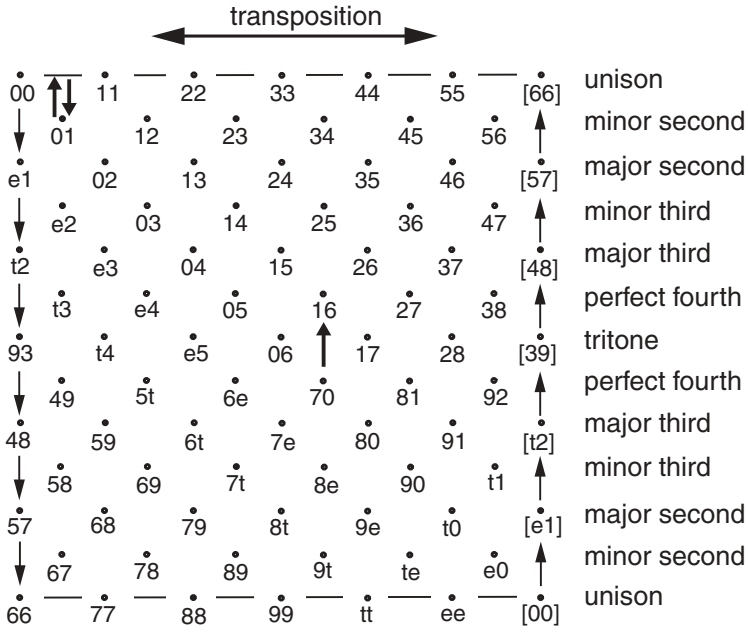


Fig. 2. The orbifold $\mathbb{T}^2/\mathcal{S}_2$. C = 0, C♯ = 1, etc., with B♭ = t, and B = e. The left edge is given a half twist and identified with the right. The voice leadings (C, D♭) → (D♭, C) and (C, G) → (C♯, F♯) are shown; the first reflects off the singular boundary.

Table 1. Common sonorities in Western tonal music. The center column lists the best equal-tempered approximation to the first n pitch classes of the harmonic series; the right column lists other good approximations. All divide the octave evenly or nearly evenly.

Size	Best approximation	Other approximations
2 notes	C, G	C, F♯
3 notes	C, E, G	C, E♭, G♭
		C, E♭, G
		C, E, G♯
4 notes	C, E, G, B♭	C, E♭, G♭, A
		C, E♭, G♭, B♭
		C, E♭, G, B♭
		C, E, G, B
5 notes	C, D, E, G, B♭	C, D, E, G, A
		C, D, E, G, B
6 notes	C, D, E, F♯, G, B♭	C, D, E♭, F, G, B♭
		C, D, E, F♯, G♯, B♭
7 notes	C, D, E, F♯, G, A, B♭	C, D, E, F, G, A, B♭
		C, D, E♭, F♯, G, A, B♭

voice leadings. Traditional tonal music exploits this possibility (Fig. 1, A to C, and movie S4). This central feature of Western counterpoint is made possible by composers' interest in the harmonic property of acoustic consonance.

A chord with duplicate pitch classes is permutationally symmetrical (P-symmetrical) because there is some nontrivial permutation of its notes that is a trivial voice leading. These chords lie on the singular boundaries of the orbifolds. Nearly P-symmetrical chords, such as {E, F, G♭}, are near these chords and contain several notes that are clustered close together. Efficient voice leadings permuting the clustered notes bounce off the nearby boundaries (Fig. 2 and movies S2 and S4). Such voice leadings can be independent and nontrivial. Trivial voice leadings are musically inert; therefore, as with T-symmetry, composers have reason to prefer near P-symmetry to exact P-symmetry.

Nearly P-symmetrical chords such as {B, C, D♭} are considered to be extremely dissonant. They are well-suited to static music in which voices move by small distances within an unchanging harmony (Fig. 1D). Such practices are characteristic of recent atonal composition, particularly the music of Ligeti and Lutoslawski. From the present perspective, these avant-garde techniques are closely related to those of traditional tonality: They exploit one of three fundamental symmetries permitting efficient, independent voice leading between transpositionally or inversionally related chords.

A chord is inversionally symmetrical (I-symmetrical) if it is invariant under reflection in pitch-class space. Nearly I-symmetrical chords are near these chords and can be found throughout the orbifolds (*16*). For example, the F♯ half-diminished seventh chord {6, 9, 0, 4} and the F dominant seventh chord {5, 9, 0, 3} are related by inversion and are very close to the I-symmetrical chord {5.5, 9, 0, 3.5}. Consequently, we can find an efficient voice leading between them, (6, 9, 0, 4) → (5, 9, 0, 3) (Fig. 1C) (*16*). Nearly T-symmetrical chords, such as the C major triad, and nearly P-symmetrical chords, such as {C, D♭, E♭}, can also be nearly I-symmetrical. Consequently, I-symmetry is exploited in both tonal and atonal music. It plays a salient role in the 19th century, particularly in the music of Schubert (*22*), Wagner (*23*), and Debussy (Fig. 1C).

The preceding ideas can be extended in several directions. First, one might examine in detail how composers have exploited the geometry of musical chords. Second, one could generalize the geometrical approach by considering quotient spaces that identify transpositionally and inversionally related chords (*24*). Third, because cyclical rhythmic patterns can also be modeled as points on $\mathbb{T}^n/\mathcal{S}_n$, one could use these spaces to study African and other non-Western rhythms. Fourth, one could investigate how distances in the orbifolds relate to perceptual judgments of chord similarity. Finally, understanding the relation between harmony and counterpoint may suggest new techniques to contemporary composers.

References and Notes

1. C. Masson, *Nouveau Traité des Règles pour la Composition de la Musique* (Da Capo, New York, 1967).
2. O. Hostinský, *Die Lehre von den musikalischen Klängen* (H. Dominicus, Prague, 1879).
3. D. Huron, *Mus. Percept.* **19**, 1 (2001).
4. J. D. Heinichen, *Der General-Bass in der Composition* (G. Olms, New York, 1969).
5. R. Cohn, *J. Mus. Theory* **41**, 1 (1997).
6. J. Roeder, thesis, Yale University (1984).
7. E. Agmon, *Musikometrica* **3**, 15 (1991).
8. R. Cohn, *Mus. Anal.* **15**, 9 (1996).
9. C. Callender, *Mus. Theory Online* **10** (2004) (http://mto.societymusictheory.org/issues/mto.04.10.3/mto.04.10.3.callender.pdf).
10. G. Mazzola, *The Topos of Music* (Birkhäuser, Boston, 2002).
11. R. Morris, *Mus. Theory Spectrum* **20**, 175 (1998).
12. J. Douthett, P. Steinbach, *J. Mus. Theory* **42**, 241 (1998).
13. J. Straus, *Mus. Theory Spectrum* **25**, 305 (2003).
14. F. Attneave, R. Olson, *Am. Psychol.* **84**, 147 (1971).
15. R. Shepard, *Psychol. Rev.* **89**, 305 (1982).
16. See supporting material on *Science* Online.
17. I. Satake, *Proc. Natl. Acad. Sci. U.S.A.* **42**, 359 (1956).
18. W. Thurston, *The Geometry and Topology of Three-Manifolds* (www.msri.org/publications/books/gt3m).
19. ChordGeometries 1.1 (http://music.princeton.edu/~dmitri/ChordGeometries.html).
20. R. Cohn, *J. Mus. Theory* **35**, 1 (1991).
21. W. Sethares, *Tuning, Timbre, Spectrum, Scale* (Springer, New York, 2005).
22. R. Cohn, *19th Cent. Mus.* **22**, 213 (1999).
23. B. Boretz, *Perspect. New Mus.* **11**, 146 (1972).
24. C. Callender, I. Quinn, D. Tymoczko, paper presented at the John Clough Memorial Conference, University of Chicago, 9 July 2005.

25. Thanks to D. Biss, C. Callender, E. Camp, K. Easwaran, N. Elkies, P. Godfrey Smith, R. Hall, A. Healy, I. Quinn, N. Weiner, and M. Weisberg.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/72/DC1
Materials and Methods
Figs. S1 to S12
Table S1
Movies S1 to S4
Soundfile S1
References

15 February 2006; accepted 26 May 2006
10.1126/science.1126287

A High-Brightness Source of Narrowband, Identical-Photon Pairs

James K. Thompson,[1]* Jonathan Simon,[2] Huanqian Loh,[1] Vladan Vuletić[1]

We generated narrowband pairs of nearly identical photons at a rate of 5 × 10⁴ pairs per second from a laser-cooled atomic ensemble inside an optical cavity. A two-photon interference experiment demonstrated that the photons could be made 90% indistinguishable, a key requirement for quantum information-processing protocols. Used as a conditional single-photon source, the system operated near the fundamental limits on recovery efficiency (57%), Fourier transform–limited bandwidth, and pair-generation-rate–limited suppression of two-photon events (factor of 33 below the Poisson limit). Each photon had a spectral width of 1.1 megahertz, ideal for interacting with atomic ensembles that form the basis of proposed quantum memories and logic.

The generation of photon pairs is useful for a broad range of applications, from the fundamental [exclusion of hidden-variable formulations of quantum mechanics (*1*)] to the more practical [quantum cryptography (*2*) and quantum computation (*3*)]. A key parameter determining the usefulness of a particular source is its brightness, i.e., how many photon pairs per second are generated into a particular electromagnetic mode and frequency bandwidth. Parametric downconverters based on nonlinear crystals are excellent sources of photon pairs, but they are comparatively dim because their photon bandwidths range up to hundreds of GHz. However, new applications are emerging that demand large pair-generation rates into the narrow bandwidths (5 MHz) suitable for strong interaction of the photons with atoms and molecules (*2*, *4*–*7*).

We report the development of a source of photon pairs with spectral brightness near fundamental physical limitations and approximately three orders of magnitude greater than the best current devices based on nonlinear crystals (*8*). Unlike parametric downconverters, however, the atomic ensemble can additionally act as a quantum memory and store the second photon, allowing triggered (i.e., deterministic) generation of the second photon. Triggered delays of up to 20 μs have been demonstrated (*9*–*15*), and it is expected that optical lattices hold the potential to extend the lifetime of these quantum memories to seconds (*9*). Lastly, proposed applications in quantum information (*2*, *3*) rely on joint measurements of single photons for which indistinguishability is crucial for high fidelity. We observe large degrees of indistinguishability in the time-resolved interference between the two generated photons (*16*–*19*).

[1]Department of Physics, MIT–Harvard Center for Ultracold Atoms, Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. [2]Department of Physics, MIT–Harvard Center for Ultracold Atoms, Harvard University, 17 Oxford Street, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed. E-mail: jkthomps@mit.edu

A range of approaches using atomic ensembles to strongly couple matter and light are actively being pursued. These include room-temperature atomic vapors (*10*, *11*) and laser-cooled atomic ensembles both in free space (*12*–*15*) and in optical cavities (*9*). Simultaneous generation of pairs of strongly correlated photons has been reported (*20*) with a 7% success rate for generation of the second photon and large violations of a Cauchy-Schwartz inequality (*21*), $G = 400 > 1$, that indicates the quantum nature of the observed correlations. Three- to fivefold suppressions of undesired two-photon events have been reported (*22*–*24*). Single photons have been captured in and released from atomic ensembles (*22*, *24*). Measurement-induced entanglement of independent ensembles of atoms has been demonstrated (*23*, *25*). The two-photon Hong-Ou-Mandel interference used here (*16*) has also been used to demonstrate the degree of indistinguishability of single photons emitted from quantum dots (*17*), and from a single atom in a high-finesse cavity (*18*, *19*).

We concentrate on the regime of minimum delay time between the generation of the photons within a pair in order to characterize the source, while keeping in mind that the present results should straightforwardly extend to the regime of delayed photon generation explored in previous work (*9*). The experimental setup consisted of a laser-cooled ensemble of $N = 10^4$ Cs atoms in the TEM_{00} mode of a low-finesse $F = 250$, single-mode optical cavity (Fig. 1 and supporting online text). Photon pairs were generated by a four-wave mixing process that relies on quantum interference in the emission from an entangled atomic ensemble (*2*) to enhance the probability of scattering a second "read" photon into the cavity to near unity given the initial scattering of a "write" photon into the cavity (Fig. 1). Without collective enhancement, the maximum probability that the read photon would be scattered into the cavity was only 7.3×10^{-4}, set by the cavity cooperativity parameter, and was nearly three orders of magnitude lower than the observed value of 0.57(9).

To first verify that the light emitted in one polarization was correlated in time with the light in the other polarization, we measured second-order correlation functions $g_{wr}(\tau)|_T$. These are simply the measured coincidence count rate between the detectors D1 and D2 normalized by the rate one would expect for two completely uncorrelated beams of the same average intensities. The label T specifies the size of the coincidence windows (and will hereafter be made implicit), and τ specifies a time offset between the write and read windows. At fixed $T = 8$ ns (Fig. 2, inset), the time-resolved cross-correlation has peak coincidence rates 100(10) times as high as those for uncorrelated beams.

To normalize out possible classical contributions to the cross-correlation data $g_{wr}(20\ \text{ns})$ of Fig. 2, we also accurately measured the auto-correlations $g_{ww}(0)$ and $g_{rr}(0)$ using two detectors for each of the write and read beams. For a bin size $T = 60$ ns, the normalized cross-correlation is $G = (g_{wr})^2/(g_{ww}g_{rr}) = 760^{+2100}_{-320}$, representing a large violation of the Cauchy-Schwarz inequality $G \leq 1$ that purely classically correlated beams must satisfy (*21*).

To assess the usefulness of the system as a source of single photons heralded by the detection of a write photon, we examined two relevant quantities: (i) the recovery efficiency, defined as the probability of generating a read photon conditioned on having detected a write photon; and (ii) the degree to which two–read photon events are suppressed below that of a Poisson distribution with the same average intensity.

A lower bound on the read recovery efficiency R is obtained from the measured detection losses q_r, combined with the measured probability of detecting a read photon given the detection of a write photon R_{det}^{cond}. The inset to Fig. 3 shows the conditional detection probability versus bin size. The integrated conditional detection probability $R_{det}^{cond} = 0.031(2)$ is estimated from the $T = 0$ intercept of a linear fit to the data at large bin size T. The read photon detection efficiency $q_r = 0.053(8)$ includes contributions from cavity mirror losses (0.45), fiber coupling (0.75), and detector quantum efficiency (0.40). Extrapolated to just outside the cavity output mirror, the recovery efficiency is $R_{cav}^{cond} = 0.26(4)$. The physical recovery efficiency for a cavity of the same linewidth, but with losses completely dominated by transmission of one of the two mirrors, is $R^{cond} = R_{det}^{cond}/q_r = 0.57(9)$. Given the low-finesse $F = 250$ of the present cavity, this ideal regime could be easily achieved with current technologies.

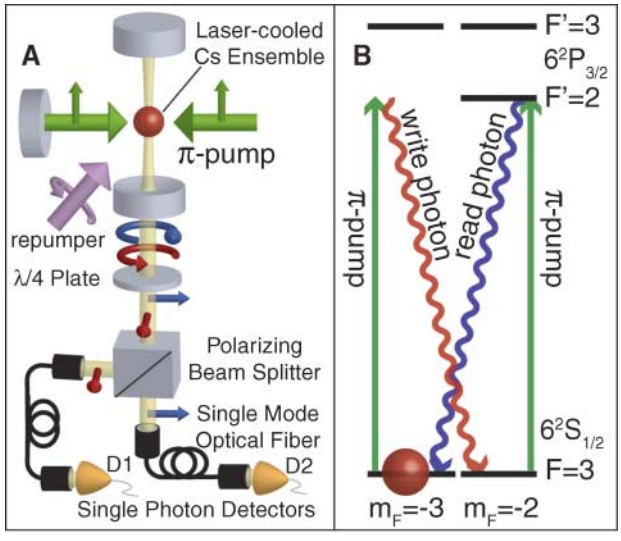


Fig. 1. (**A**) Experimental setup and (**B**) quantum states used for photon-pair generation. The tuning of the π-pump laser is chosen so that the rate of write photon scattering into the cavity is suppressed by a large detuning from resonance with any excited state, whereas the collectively stimulated generation of a read photon in the cavity proceeds rapidly via resonant coupling. This ensures that the time separation between subsequent pairs exceeds the time separation of the write and read photons within a pair—leading to large cross-correlations between the photon polarizations. The pump and emitted-photon polarizations are denoted by the smaller arrows. The π-pump in combination with a repumper (tuned to the ground $F = 4$ to excited $F' = 4$ transition) optically pumps ~95% of the atomic population into $|F = 3, m_F = -3\rangle$.

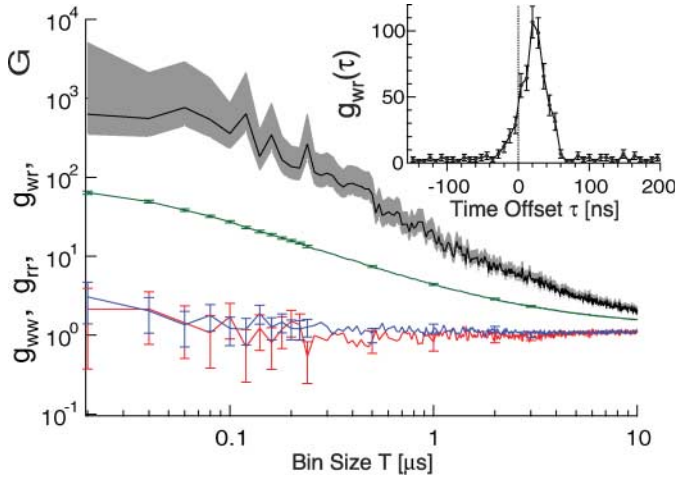


Fig. 2. Nonclassical photon-pair generation. The measured violation of the Cauchy-Schwarz inequality $G = [g_{wr}(\tau)]^2/[g_{ww}(0)g_{rr}(0)] \leq 1$ versus bin size T (black curve with 68% confidence interval), indicating large nonclassical correlations between the write and read photon beams. The inequality simply states that a cross-correlation $g_{wr}(20\ \text{ns})$ (green) arising from classical sources (i.e., pump-intensity fluctuations) must also manifest itself in the measured autocorrelations $g_{ww}(0)$ (red) and $g_{rr}(0)$ (blue). (**Inset**) A time-resolved cross-correlation function $g_{wr}(\tau)$ is shown with a peak value of 100(10).

The conditional suppression of two-photon events was measured using one detector to herald the arrival of a write photon at time t and two detectors to measure the autocorrelation R_{rr}^{cond} of the subset of the read photons that fall within a time bin of size $\pm T/2$ centered at $t + \tau$. With $\tau = 20$ ns, the conditional autocorrelation starts near unity (no suppression) at large bin sizes T (Fig. 3) due to backgrounds not correlated with the registered write photon, and it decreases monotonically until the bin size becomes comparable to the read photon emission time near $T \sim 100$ ns. The largest measured suppression of two-photon events $R_{rr}^{cond} = 0.03(3)$ occurs at $T = 60$ ns.

A conditional single-photon source can only suppress two-photon events to the fundamental limit $R_{rr}^{cond} = \varepsilon_w g_{ww}$ associated with the random emission of two pairs of photons within the same time bin T. In the ideal case, the autocorrelation of the write photons is $g_{ww} = 2$. The write photon generation probability ε_w can be extracted from the measured write photon detection probability ε^{det} via $\varepsilon_w = \varepsilon_w^{det}/q_w = 0.013(2)$ at $T = 60$ ns, thus predicting the fundamental limit for a conditional photon source at this pair-generation rate of $R_{rr}^{cond} \geq 0.026(3)$. The agreement with the measured value $R_{rr}^{cond} = 0.03(3)$ indicates that spurious background counts are not a serious limitation to the performance of the single-photon source at the present photon-generation rate.

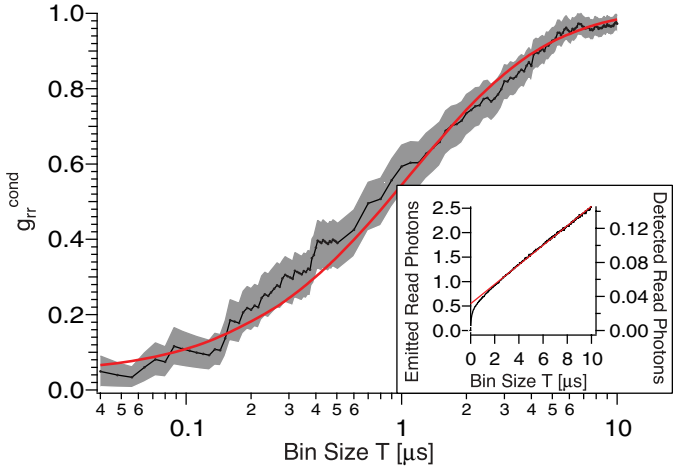


Fig. 3. Performance of the conditional (heralded) single-photon source. The fractional suppression of two-photon events g_{rr}^{cond} and (**inset**) the fractional probability of generating a read photon given the detection of a write photon are both shown versus bin size T. At large bin sizes, read photons generated by other write photons drive the conditional autocorrelation toward the classical limit of unity. As the bin size is reduced, the autocorrelation becomes highly nonclassical ($g_{rr}^{cond} < 1$). As was done for the correlation data in Fig. 2, the average estimate (not maximum-likelihood estimate) is used to avoid underestimating g_{rr}^{cond} and the 68% confidence interval (gray band) at small numbers of counts. The right and left axis of the inset show, respectively, the probabilities conditioned on the detection of a write photon for detection of a read photon (right) and the extrapolated generation of a read photon for a cavity with no mirror losses (left), 57(9)%, extrapolated to $T = 0$. The red curve in (A) is an independent prediction of g_{rr}^{cond} from combining measured background rates with the measured time dependence of the recovery efficiency.

The identicalness of the write and read photons was examined via two-photon interference at the polarizing beam splitter (Fig. 1A) (*16*–*19*). This was accomplished by analyzing the write and read photons in a polarization basis rotated by 45° with respect to the usual basis used to deterministically separate the photons. Neglecting interference between the two photons, one expects that in half the cases, the photons register a coincidence count on opposite detectors. However, if the write and read photons perfectly overlap in time and frequency, there is a complete destructive interference for the probability of a coincidence count [a so-called Hong-Ou-Mandel interference (*16*)]. The fractional reduction of the coincidence count rate below ½ of its original value is a direct measure of the degree of indistinguishability of the photons.

Figure 4 shows the measured coincidence rate expressed as a cross-correlation between detectors D_1 and D_2, when the write and read photons are polarization separated $g_{wr}(\tau)$ (Fig. 4A) and are allowed to interfere $g_{45}(\tau)$ (Fig. 4B). The destructive interference is most pronounced near $\tau = 0$, and it decreases as $|\tau|$ increases because the finite time separation allows one to infer with increasing reliability which detection event corresponds to the write photon and which to the read photon. For data sets at larger π-pump intensities for which the read photons are emitted more promptly, we observe suppressions of the two-photon coincidence rate below ½ by as much as 90(20)% integrated over $T = 5$ μs, indicating that the photons can be made nearly identical.

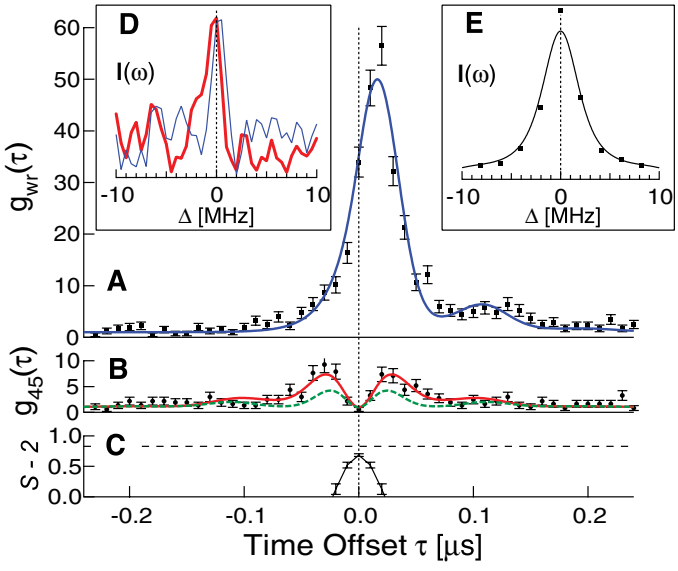


Fig. 4. Measures of identicalness and photon frequency bandwidths. (**A**) The time-resolved cross-correlation function $g_{wr}(\tau)$ and (**B**) the same function $g_{45}(\tau)$ measured in a polarization basis rotated by 45°. In the 45° basis, coincidence events are suppressed by two-photon interference resulting from the near indistinguishability of the photons. Assuming the photons have identical frequencies, the quantity $g_{45}(\tau)$ can be predicted directly from $g_{wr}(\tau)$ [green dashed curve in (B)]. The prediction is more accurate if a photon frequency difference $\Delta\omega/2\pi = 2.5$ MHz is assumed [red curve in (B)]. (**C**) The predicted violation of a Bell's inequality $S - 2 < 0$ if the photon pairs were used to produce polarization-entangled photons. The dashed line is the maximum possible violation. (**D**) The frequency bandwidths of the write (red) and read (blue) photons are determined to be 1.1(2) MHz from the displayed heterodyne beat notes. For comparison, (**E**) shows the square of the Fourier transform of $\sqrt{g_{wr}(\tau) - 1}$ taken at different parameters, indicating that the photon bandwidths are nearly transform limited.

The two-photon interference results above can be directly mapped onto a *gedanken* version

of the experiments in (*26*–*29*), wherein polarization entanglement is generated via post selection (see supporting online text and Fig. 4C). The mapping is performed assuming that quantum mechanics is correct (*1*, *27*–*29*). At $\tau = 0$, the predicted CHSH Bell's parameter is $S = 2.68(2)$, a violation of the Bell's inequality $|S| \leq 2$. The predicted violation is not closer to the theoretical maximum $S_{max} = 2\sqrt{2} \approx 2.828$ (dashed line of Fig. 4C), largely due to backgrounds set by the two-photon generation rate.

The frequency bandwidths of the write and read photons are 1.1(2) MHz, making them ideal for interacting with narrowband systems such as atoms, molecules, and optical cavities. By separately heterodyning the write and read photons with laser light derived from the π-pump laser (measured linewidth of 50 kHz), we obtained the power spectral density of the photons from the Fourier transform of the measured second-order autocorrelation function (Fig. 4D). The photons are nearly Fourier-transform limited, as can be seen from the 2-MHz full width at half-maximum power spectrum (Fig. 4E) of the measured cross-correlation function $g_{wr}(\tau)$ taken at slightly different parameters.

These measurements show that pairs of nearly identical photons are generated at an approximate rate of 5×10^4 pairs/s into a single Gaussian transverse mode. The spectral brightness of 5×10^4 pairs/s per MHz^{-1} is $\sim 10^3$ times as bright as the best sources based on parametric downconversion with nonlinear crystals (*8*). The system can operate very near fundamental limits on recovery efficiency, photon bandwidth, and two-photon suppression for a conditional single-photon source. In addition, identical photon pairs are necessary for certain quantum information protocols such as quantum computation with linear optics (*3*). The identical photon pairs also have potential applications for sub-shotnoise spectroscopy of atomic ensembles.

References and Notes

1. A. Aspect, P. Grangier, G. Roger, *Phys. Rev. Lett.* **49**, 91 (1982).
2. L.-M. Duan, M. D. Lukin, J. I. Cirac, P. Zoller, *Nature* **414**, 413 (2001).
3. E. Knill, R. Laflamme, G. Milburn, *Nature* **409**, 46 (2001).
4. Q. A. Turchette, C. J. Hood, W. Lange, H. Mabuchi, H. J. Kimble, *Phys. Rev. Lett.* **75**, 4710 (1995).
5. D. F. Phillips, A. Fleischhauer, A. Mair, R. L. Walsworth, M. D. Lukin, *Phys. Rev. Lett.* **86**, 783 (2001).
6. A. André, L.-M. Duan, M. D. Lukin, *Phys. Rev. Lett.* **88**, 243602 (2002).
7. B. Julsgaard, A. Kozhekin, E. S. Polzik, *Nature* **413**, 400 (2001).
8. F. Konig, E. J. Mason, F. N. C. Wong, M. A. Albota, *Phys. Rev. A.* **71**, 033805 (2005).
9. A. T. Black, J. K. Thompson, V. Vuletić, *Phys. Rev. Lett.* **95**, 133601 (2005).
10. C. H. van der Wal *et al.*, *Science* **301**, 196 (2003).
11. M. D. Eisaman *et al.*, *Phys. Rev. Lett.* **93**, 233602 (2004).
12. A. Kuzmich *et al.*, *Nature* **423**, 731 (2003).
13. C. W. Chou, S. V. Polyakov, A. Kuzmich, H. J. Kimble, *Phys. Rev. Lett.* **92**, 213601 (2004).
14. S. V. Polyakov, C. W. Chou, D. Felinto, H. J. Kimble, *Phys. Rev. Lett.* **93**, 263601 (2004).
15. D. Matsukevich, A. Kuzmich, *Science* **306**, 663 (2004).
16. C. K. Hong, Z. Y. Ou, L. Mandel, *Phys. Rev. Lett.* **59**, 2044 (1987).
17. C. Santori, D. Fattal, J. Vuckovic, G. S. Solomon, Y. Yamamoto, *Nature* **419**, 594 (2002).
18. T. Legero, T. Wilk, M. Hennrich, G. Rempe, A. Kuhn, *Phys. Rev. Lett.* **93**, 070503 (2004).
19. T. Legero, T. Wilk, A. Kuhn, G. Rempe, *Appl. Phys. B* **77**, 797 (2003).
20. V. Balić, D. A. Braje, P. Kolchin, G. Y. Yin, S. E. Harris, *Phys. Rev. Lett.* **94**, 183601 (2005).
21. J. F. Clauser, *Phys. Rev. D* **9**, 853 (1974).
22. M. D. Eisaman *et al.*, *Nature* **438**, 837 (2005).
23. C. W. Chou *et al.*, *Nature* **438**, 828 (2005).
24. T. Chaneliere *et al.*, *Nature* **438**, 833 (2005).
25. D. N. Matsukevich *et al.*, *Phys. Rev. Lett.* **96**, 030405 (2006).
26. J. F. Clauser, M. A. Horne, A. Shimony, R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
27. Y. H. Shih, C. O. Alley, *Phys. Rev. Lett.* **61**, 2921 (1988).
28. Z. Y. Ou, L. Mandel, *Phys. Rev. Lett.* **61**, 50 (1988).
29. D. Fattal *et al.*, *Phys. Rev. Lett.* **92**, 037903 (2004).

30. This work was supported in parts by grants from NSF, Defense Advanced Research Projects Agency, Army Research Office, and the Sloan Foundation. J.S. acknowledges a National Defense Science and Engineering Graduate Fellowship from the U.S. Department of Defense. H.L. acknowledges a Josephine de Karman Fellowship and a A*STAR (Singapore) Scholarship.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/74/DC1
SOM Text
References

21 March 2006; accepted 1 June 2006
10.1126/science.1127676

Surface Crystallization in a Liquid AuSi Alloy

Oleg G. Shpyrko,[1,2*] Reinhard Streitel,[1] Venkatachalapathy S. K. Balagurusamy,[1] Alexei Y. Grigoriev,[1] Moshe Deutsch,[3] Benjamin M. Ocko,[4] Mati Meron,[5] Binhua Lin,[5] Peter S. Pershan[1]

X-ray measurements reveal a crystalline monolayer at the surface of the eutectic liquid $Au_{82}Si_{18}$, at temperatures above the alloy's melting point. Surface-induced atomic layering, the hallmark of liquid metals, is also found below the crystalline monolayer. The layering depth, however, is threefold greater than that of all liquid metals studied to date. The crystallinity of the surface monolayer is notable, considering that AuSi does not form stable bulk crystalline phases at any concentration and temperature and that no crystalline surface phase has been detected thus far in any pure liquid metal or nondilute alloy. These results are discussed in relation to recently suggested models of amorphous alloys.

Surface melting—the coexistence of a liquid surface layer with the bulk crystal at temperatures below the bulk melting point T_m—has been observed in a wide range of materials (*1*, *2*) and occurs because the entropy of molecules at the free surface is greater than that in the bulk because of the reduced number of their near neighbors. The opposite effect, surface freezing, where a crystalline surface layer coexists with its molten bulk, is much rarer. Surface freezing has been observed, however, in complex liquids composed of high-anisotropy molecules, such as molten unary or binary alkanes and their derivatives (*3*), and in liquid crystals (*4*). Theory assigns the occurrence of this effect to the highly anisotropic shape of the molecules and to their lengths being greater than the interfacial width (*5*).

Freezing of the surface-segregated component into a two-dimensional (2D) solid layer has also been reported recently in the very dilute binary metallic alloys $Ga_{99.948}Pb_{0.052}$ (*6*) and $Ga_{99.986}Tl_{0.014}$ (*7*). A different ordering effect, surface-induced layering consisting of stratified layers (Fig. 1) near the vapor interface (*8*, *9*), has been observed in all liquid metals and alloys studied to date. The decay of the layering order with depth is exponential and has a range equal to the bulk liquid correlation length (two to three atomic diameters). No surface-parallel ordering was found within these layers in any elemental liquid metal. Similar layering, along with epitaxially induced surface-parallel order, has also been observed in both metallic and nonmetallic liquids near solid/liquid interfaces (*10*–*12*).

We present evidence for surface crystallization and enhanced surface layering in the liquid $Au_{82}Si_{18}$ eutectic alloy of a type unlike that previously reported for any liquid metal or alloy. A surface monolayer that exhibits lateral long-range crystalline order was found above the eutectic temperature $T_e = 359°C$. Beneath this monolayer, seven to eight layers occur that are liquid in the lateral direction but well defined in the normal direction. The crystalline surface monolayer and the enhancement of the

[1]Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [2]Center for Nanoscale Materials, Argonne National Laboratory, Argonne, IL 60439, USA. [3]Department of Physics, Bar-Ilan University, Ramat-Gan 52900, Israel. [4]Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, NY 11973, USA. [5]Center for Advanced Radiation Sources, University of Chicago, Chicago, IL 60637, USA.

*To whom correspondence should be addressed. E-mail: oshpyrko@anl.gov

surface-induced layering range beyond the two to three layers observed in all other liquid metals studied to date clearly have a common origin. The surface-frozen monolayer undergoes a first-order transition into a different surface phase 12°C above T_e.

These unusual surface structures probably result from the equally unusual bonding properties of metastable amorphous bulk AuSi. AuSi, the first metallic alloy found to exhibit a glassy solid phase (*13*), remains one of the most puzzling amorphous solids. Silicon-rich amorphous AuSi is a semiconductor, has a low packing density, and has a low atomic coordination number (4 to 5). Its main structural motif is a continuous random network of covalently bonded Si atoms. The Au-rich alloy, however, is a metallic glass, almost as dense as a face-centered cubic lattice; has a high atomic coordination number (8 to 9); and has a random hard-sphere packing (*14*). Such random packing in amorphous metals was recently shown to consist of interpenetrating clusters, the outer atoms of which are shared by adjacent clusters (*15*, *16*). Because Si has a lower surface tension than Au, the surface of the liquid AuSi alloy is Si-rich, so atomic packing and bonding at the surface might be expected to be more like that of the covalently bonded Si-rich alloys than the metallic Au-rich bulk.

The existence of a very deep eutectic (for $Au_{82}Si_{18}$) at $T_e \approx 359$°C, much below T_m = 1063°C of Au and T_m = 1412°C of Si (Fig. 2A, inset), arises from the bonding effects discussed above. Below T_e, AuSi phase-separates in thermodynamic equilibrium into crystalline Au and Si, with no mutual solid solubility and no stable crystalline intermetallic compounds, whereas metastable amorphous AuSi phases can be achieved by rapid quenching, sputtering, and other techniques (*14*).

X-ray reflectivity off a liquid surface, $R(q_z)$, is measured as a function of the grazing angle of incidence α. Here $q_z = (4\pi/\lambda)\sin\alpha$ is the surface-normal wave vector transfer and λ is the x-rays' wavelength. The ratio R/R_F, where R_F is the theoretical reflectivity off of an ideally flat and abrupt liquid-vapor interface, depends on the surface-normal electron density profile ρ(*z*). A layered interface produces a Bragg-like peak in $R(q_z)/R_F(q_z)$ because of the constructive interference of the rays reflected from the periodically ordered surface layers (*8*). The larger the number of layers, the higher is the layering peak. Figure 2A demonstrates that the AuSi layering peak at 370°C is at least one order of magnitude higher than for the standard layering profile observed in all elemental liquids measured to date (*17*). This result implies that there are more than the two to three layers found in all previously measured liquid metals.

Indeed, a theoretical model fit to the measured values of $R(q_z)/R_F(q_z)$ (Fig. 2A, red line) yields the ρ(*z*) curve shown in Fig. 2B. Although the finer details of this ρ(*z*) curve may not be unique, two features were independent of the model used: the 2.5-nm-thick layering range (seven to eight well-defined atomic layers) and the Si enrichment of the top layer, indicated by a ρ value less than that of ρ_{bulk}, which corresponds to ~70 atomic % Si. This value agrees well with the 67 atomic % calculated from the Gibbs adsorption rule for an ideal binary solution (*18*).

No variation was found in the measured $R(q_z)$ from T_e = 359°C up to 371°C. At 371°C, $R(q_z)/R_F(q_z)$ changed abruptly (Fig. 2A, curve with black squares). By monitoring $R(q_z)$ at a fixed q_z while varying the temperature (Fig. 2B, inset), we found the surface phase transition at 371°C to be reproducible and to exhibit no hysteresis (<0.1°C). The narrow width of 0.17°C suggests a first-order phase transition.

The surface-parallel structure was explored by grazing incidence x-ray diffraction (GIXD). X-rays impinging on the liquid AuSi surface well below the critical angle penetrated the surface only evanescently, to a depth of ~1.4 nm (*19*), and produced a diffraction pattern for only the top ~5 atomic surface layers. The GIXD pattern measured for 359°C ≤ *T* ≤ 371°C (Fig. 3) showed sharp diffraction peaks indicative of long-range lateral ordering. A broad peak, characteristic of a liquid, was also observed. The GIXD pattern was indexed in a 2D rectangu-

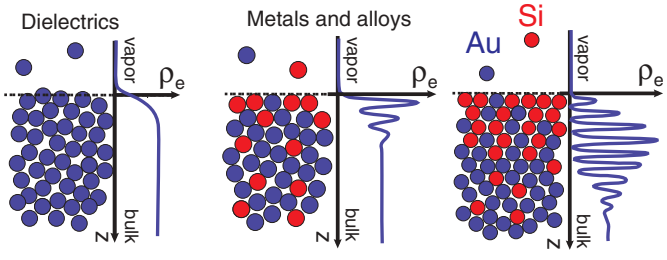


Fig. 1. Typical atomic surface structure and corresponding electron density profiles $\rho_e(z)$ of nonlayered dielectric liquids (**left**), standard layering in liquid metals and alloys (**middle**), and enhanced layering in AuSi (**right**).

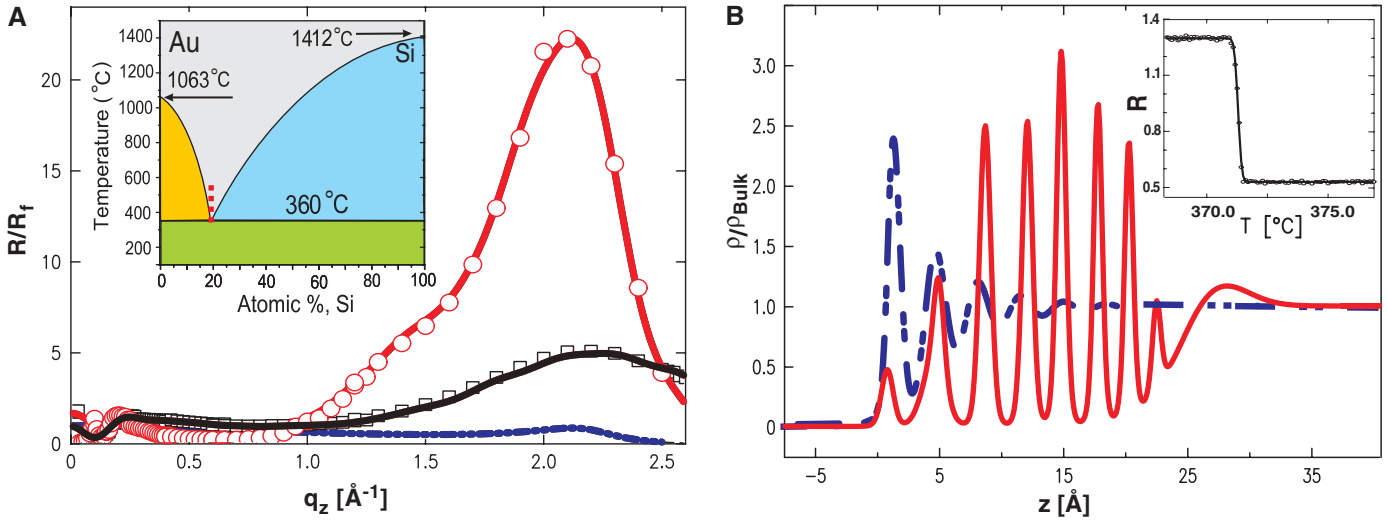


Fig. 2. (**A**) Fresnel-normalized x-ray reflectivity at 370°C (circles) and 375°C (squares), the corresponding model fits (red and black lines, respectively), and the curve expected for standard layering (dashed blue line). (Inset) Bulk phase diagram of AuSi. The gray area represents a liquid mixture phase, and the blue and yellow areas indicate the phase coexistence of solid Si or solid Au with a liquid alloy. The green area corresponds to phase-separated solid Au and solid Si. (**B**) Surface-normal electron density profiles corresponding to the same-line models in (A). (Inset) Reflectivity at fixed q_z = 1.0 Å^{-1} (circles) versus temperature, with a fit (black line) by an error function centered at 371.29°C with a width of 0.17°C.

lar lattice of dimensions a = 7.386 Å and b = 9.386 Å. A full-pattern refinement yielded the Au_4Si_8 structure shown in the right inset of Fig. 3. The high-T surface phase, which forms at 371°C and exists up to at least 410°C, also exhibits a set of sharp GIXD peaks but at different q_{xy} positions (*20*) than those of the 359°C < T < 371°C surface phase.

The GIXD peak intensities in the low-T phase were not affected by sample rotation around the surface-normal axis, indicating that the diffracting monolayer consists of a fine powder of randomly oriented crystallites. Debye-Scherrer analysis of the line shapes, measured with a high-resolution analyzer crystal, yields a typical crystallite size of ~ 2 to 10 μm. The measured Bragg rods (see the supporting online material) are surface-normal, indicating a quasi-2D crystalline structure. The ~1.5 Å^{-1} width of the rod's q_z intensity distribution implies a crystalline layer thickness of d = π/(1.5 Å^{-1}), where d is approximately equal to 2 Å (that is, a monolayer). This result agrees with the d value of ~2.5 Å that is estimated from the ratio of the integrated intensities of the GIXD peaks and the broad liquid peak.

Notably, despite the crystalline order, the capillary surface dynamics are still liquid-like; the diffuse scattering measured away from the specular peak exhibits a power-law behavior characteristic of the height-height correlations in liquid surfaces (*21*) (Fig. 3, left inset). The line shape of the diffuse scan (solid line in left inset of Fig. 3) is well reproduced by the capillary wave theory profiles if we use γ_{AuSi} = 780 mN/m convolved with the experimental resolution function. The agreement with capillary wave theory also indicates that in spite of the larger layering depth, the subsurface ordered layers are laterally liquid, as is the case for all previously measured liquid metals.

The crystalline AuSi monolayer's structure does not resemble those of pure Au, pure Si, or any of the reported metastable bulk intermetallics. However, crystalline phases with unit cell dimensions a = 7.44 Å and b = 9.33 Å, similar to those observed in our study, were reported in thin Au films deposited on a Si(111) surface (*22*). Some of these phases were thought to be surface phases, and some may exist also in the bulk. No reliable thicknesses of these phases could be obtained by the low-energy electron diffraction and Auger electron spectroscopy techniques that were used. A clear understanding of the formation mechanism, stability, and surface specificity of the crystalline phases in this system has not yet emerged (*23*). However, the strong Si-Si bonding (*24*) was suggested to play a crucial role, as well as the quasi-2D atomic coordination near the surface (*25*). In our case, the Gibbs adsorption surface enrichment by Si atoms is likely to facilitate the chemical Au-Si bonds that stabilize the crystalline surface layer but that evolve to metallic-like bonding away from the surface (*26*).

The Si packing in the crystalline surface monolayer (Fig. 3, right inset) resembles the covalently bonded, network-forming Si chains suggested to stabilize the amorphous structure of Si-rich AuSi alloys (*15*, *16*). A comparison of the unit cell structure presented in Fig. 3 with figure 1A of Miracle (*15*) indicates that the unit cell packing here is close to the interpenetrating cluster structure recently suggested as the main structural motif of bulk amorphous solids.

The formation of a surface-frozen phase is typically marked by a change in the slope of the surface tension versus temperature curve, $\gamma(T)$, from $d\gamma/dT > 0$ for a crystalline surface to $d\gamma/dT < 0$ for a liquid surface (*3*). Indeed, a positive slope is reported for AuSi above melting (*27*), in line with the crystalline phase reported here. The change to a negative slope, reported to occur at $T \approx$ 800° to 900°C (*27*), should indicate the melting of the ordered surface structure. Similar positive $\gamma(T)$ slopes above melting have been recently reported for AgSn, AgBi, AgIn, InCu, CuSn, MnSn, and AuZn binary alloys, implying that the formation of surface-frozen phases may not be an entirely uncommon phenomenon in multicomponent liquid alloys.

Beyond its importance for understanding the physics underlying amorphous metallic alloys, AuSi is also of high technological importance, because Au is widely used in interconnecting integrated circuits on Si substrates. AuSi also has important nanoscale applications such as the self-assembly of Si nanowires (*28*) and low-temperature bonding in micro- and nanoelectromechanical devices (*29*). Surface phases are of particular importance for nanotechnology, because properties of objects at the nanometer scale are expected to be dominated by surfaces and interfaces. The discovery of previously unidentified structures bridging the gap between 2D and 3D phases is expected, therefore, to have far-reaching consequences for both fundamental and applied research.

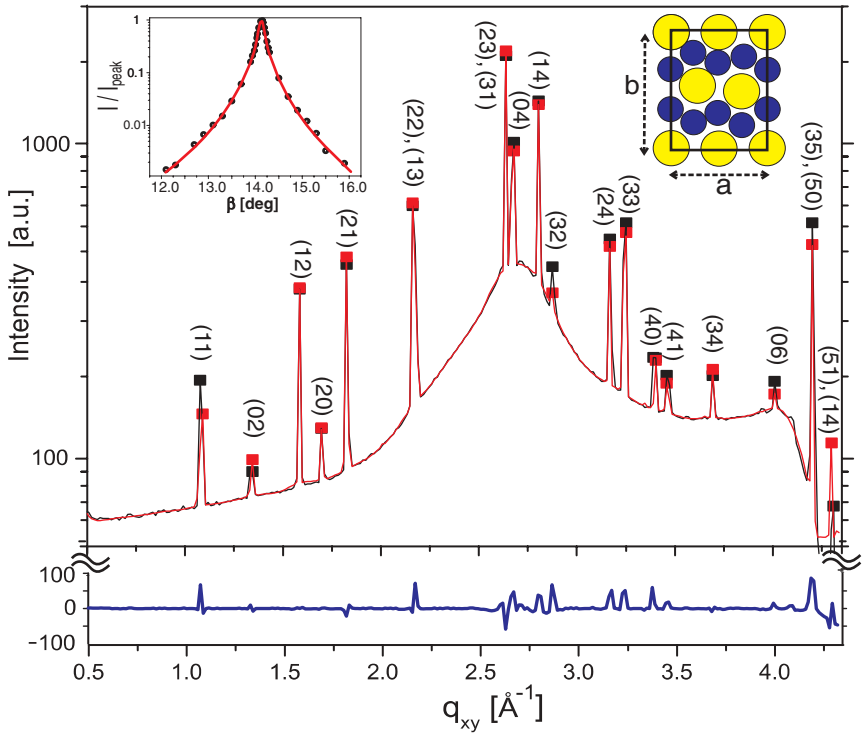


Fig. 3. Measured GIXD pattern (black line), theoretical fit (red line), and their difference (blue line) for the 359°C < T < 371°C surface phase. a.u., arbitrary units. (**Left inset**) Diffuse scattering profile versus the output detector angle β for a fixed incidence angle α = 14.11° and its fit (line) by the capillary wave theory prediction for γ = 780 mN/m. I/I_{peak} is the x-ray intensity normalized by the scan's peak intensity value. (**Right inset**) Crystal unit cell obtained from the GIXD pattern, where a = 7.386 Å and b = 9.386 Å. Au, yellow circles; Si, blue circles.

References and Notes

1. J. W. M. Frenken, J. F. Van der Veen, *Phys. Rev. Lett.* **54**, 134 (1985).
2. J. G. Dash, *Phys. Today* **38**, 26 (1985).
3. X. Z. Wu *et al.*, *Science* **261**, 1018 (1993).
4. Z. Dogic, *Phys. Rev. Lett.* **91**, 165701 (2003).
5. A. V. Tkachenko, Y. Rabin, *Phys. Rev. Lett.* **76**, 2527 (1996).
6. B. Yang, D. Gidalevitz, D. Li, Z.-q. Huang, S. A. Rice, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13009 (1999).
7. B. Yang, D. Li, S. A. Rice, *Phys. Rev. B* **67**, 212103 (2003).
8. O. M. Magnussen *et al.*, *Phys. Rev. Lett.* **74**, 4444 (1995).
9. M. J. Regan *et al.*, *Phys. Rev. Lett.* **75**, 2498 (1995).
10. S. E. Donnelly *et al.*, *Science* **296**, 507 (2002).
11. S. H. Oh, Y. Kauffmann, C. Scheu, W. D. Kaplan, M. Rühle, *Science* **310**, 661 (2005).
12. W. J. Huisman *et al.*, *Nature* **390**, 379 (1997).
13. W. Klement Jr., R. H. Willens, P. Duwez, *Nature* **187**, 869 (1960).

14. J. Weissmüller, *J. Non-Cryst. Solids* **142**, 70 (1992).
15. D. B. Miracle, *Nat. Mater.* **3**, 697 (2004).
16. H. W. Sheng *et al.*, *Nature* **439**, 419 (2006).
17. Our unpublished measurements on the liquid eutectic alloys of AuSn and AuGe yielded layering peaks 30 to 50 times lower than that of AuSi.
18. O. G. Shpyrko *et al.*, *Phys. Rev. Lett.* **95**, 106103 (2005).
19. Details of the GIXD experiment are available on *Science* Online.
20. Because of insufficient reproducibility, the discussion of this pattern must be deferred.
21. S. K. Sinha, E. B. Sirota, S. Garoff, H. B. Stanley, *Phys. Rev. B* **38**, 2297 (1988).
22. A. K. Green, E. Bauer, *J. Appl. Phys.* **47**, 1284 (1976).
23. J. F. Chang *et al.*, *Mat. Chem. Phys.* **83**, 199 (2004).
24. H. S. Chen, D. Turnbull, *J. Appl. Phys.* **38**, 3646 (1967).
25. S. L. Molodtsov, C. Laubschat, G. Kaindl, A. M. Shikin, V. K. Adamchuk, *Phys. Rev. B* **44**, 8850 (1991).
26. V. A. Filonenko, *Russ. J. Phys. Chem.* **43**, 874 (1969).
27. L. Bischoff, J. Teichert, T. Ganetsos, G. L. R. Mair, *J. Phys. D Appl. Phys.* **33**, 692 (2000).
28. J. B. Hannon, S. Kodambaka, F. M. Ross, R. M. Tromp, *Nature* **440**, 69 (2006).
29. Y. T. Cheng, L. W. Lin, K. Najafi, *J. Microelectromech. Syst.* **9**, 3 (2000).
30. This work was supported by the U.S. Department of Energy (DOE) grant DE-FG02-88-ER45379 and the U.S.-Israel Binational Science Foundation, Jerusalem. Brookhaven National Laboratory is supported by U.S. DOE contract DE-AC02-98CH10886. We acknowledge beamline assistance from J. Gebhardt, T. Graber, and H. Brewer at the Chemistry and Materials Science Sector of the Center for Advanced Radiation Sources (ChemMatCARS). ChemMatCARS Sector 15 is principally supported by NSF/DOE grant CHE0087817. The Advanced Photon Source is supported by U.S. DOE contract W-31-109-Eng-38.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/77/DC1
Materials and Methods
Fig. S1
References and Notes

4 April 2006; accepted 25 May 2006
10.1126/science.1128314

Probing the Solvent-Assisted Nucleation Pathway in Chemical Self-Assembly

Pascal Jonkheijm,[1]* Paul van der Schoot,[2] Albertus P. H. J. Schenning,[1]† E. W. Meijer[1]†

Hierarchical self-assembly offers a powerful strategy for producing molecular nanostructures. Although widely used, the mechanistic details of self-assembly processes are poorly understood. We spectroscopically monitored a nucleation process in the self-assembly of π-conjugated molecules into helical supramolecular fibrillar structures. The data support a nucleation-growth pathway that gives rise to a remarkably high degree of cooperativity. Furthermore, we characterize a helical transition in the nucleating species before growth. The self-assembly process depends strongly on solvent structure, suggesting that an organized shell of solvent molecules plays an explicit role in rigidifying the aggregates and guiding them toward further assembly into bundles and/or gels.

Chemical self-assembly offers a very attractive approach for constructing complex, supramolecular nanostructures. Hierarchical processes, typical of chemical self-assembly, spontaneously produce ordered ensembles of single or multiple molecular components and are ubiquitous in chemistry, physics, materials science, and biology (*1*). For example, a large variety of molecules has been reported that form gels through three-dimensional networks of (bundles of) fibers (*2*). Highly versatile biomaterials have been produced by self-assembly of peptide amphiphiles (*3*), and (semi)conducting tubes and rods have been achieved from properly chosen π-conjugated oligomeric building blocks (*4*). In many cases, the resulting fibrillar structures are helical, and a preferred handedness is obtained by the introduction of stereocenters into the building blocks (*5*). Ensuring that the components aggregate in a specific motif, however, remains a formidable task; molecular components are easily trapped in kinetically stable arrangements of varying topology (*6*). The application of spectroscopy in combination with scattering and microscopy techniques has provided a reasonable view of the final self-assembled structures, yet a thorough understanding of the processes leading up to these structures remains elusive. Unravelling such structural pathways is crucial for the widely sought goal of extending rational synthesis into the nanoscale regime (*7*).

Although the growth of fibrillar structures typically requires nucleation, prior studies have not focused on the nature and properties of the nuclei. This lack of data is in sharp contrast to the abundant studies on the crystallization of small molecules (*8*) or polymers (*9*), and on the aggregation of proteins (*10*). For instance, notable insight into the formation of actin filaments, microtubules, and viral capsids by a nucleation-growth mechanism was obtained after the pioneering work of Klug (*11*) and Caspar (*12*). More recently, this insight was used to study pathologies stemming from amyloid deposits and diseases related to protein aggregation (*13*). Phenomenologically, the different phases of protein aggregation should be very similar to those involved in chemical self-assembly. Two models are commonly invoked to describe protein aggregation. Isodesmic self-assembly (*10*), also called multistage open association or ladder- or free-association model, is noncooperative, and the association constant is independent of the size of the object. Nucleated self-assembly, also called nucleation growth or initiation elongation, is characterized by a size-dependent association constant that gives rise to cooperative kinetics. We have sought to characterize such processes in chemical synthetic systems.

The fibrillar structures we examined comprise oligo(*p*-phenylenevinylene) derivatives (OPV-*x*) with chiral side chains capped on one end by a tridodecyloxybenzene and on the other by a ureidotriazine tailored for self-complementary fourfold hydrogen bonding (structure shown in Fig. 1) (*14*, *15*). Previous

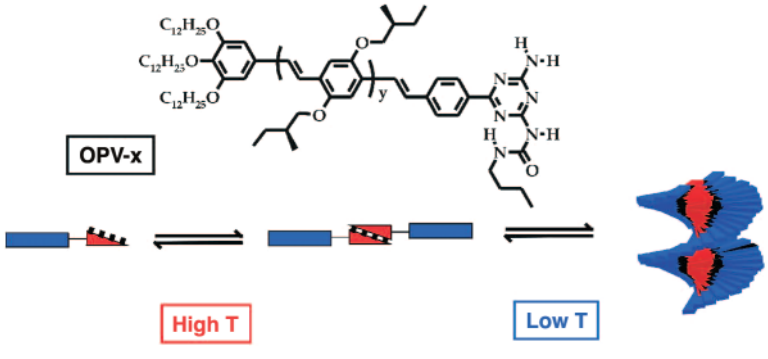


Fig. 1. Molecular structure of the oligo(*p*-phenylenevinylene) derivatives OPV-*x* [for OPV-3, *y* (number of dialkoxybenzene units) = 1; OPV-4, *y* = 2; and OPV-5, *y* = 3] and schematic representation of the self-assembly process, with blue blocks representing the OPV backbone and red wedges representing the hydrogen-bonding end groups.

[1]Laboratory of Macromolecular and Organic Chemistry, [2]Polymer Physics Group, Eindhoven University of Technology, Post Office Box 513, 5600 MB Eindhoven, Netherlands.

*Present address: Chemical Biology, Max Planck Institute for Molecular Physiology, Otto Hahnstrasse 11, 44227 Dortmund, Germany.

†To whom correspondence should be addressed. E-mail: e.w.meijer@tue.nl (E.W.M.); a.p.h.j.schenning@tue.nl (A.P.H.J.S.)

temperature-dependent ultraviolet and visible absorption (UV/vis), fluorescence, and circular dichroism (CD) measurements (fig. S1) (*14*, *15*) revealed two different states of the π-conjugated chromophores in dodecane solution: discrete monomeric or hydrogen-bonded dimeric species at high temperature and helical aggregates at low temperature (Fig. 1). The hydrogen-bonded dimers have been studied in detail with scanning tunneling microscopy (STM) (*16*) and ^{1}H nuclear magnetic resonance (NMR) spectroscopy (*14*), whereas the fibrillar structural dimensions have been measured by small angle neutron scattering (SANS) and atomic force microscopy (AFM) (*15*). For example, OPV-4 exhibited a fibril persistence length of 150 nm and diameter of 5.8 nm in dodecane (table S1). The diameters of the fibrils were in close agreement with molecular modeling studies (table S1). Fluorescence microscopy revealed a perpendicular chromophore orientation within the fibrils (*17*).

By improving our spectroscopic sensitivity, we reveal in this study distinct hierarchical stages that govern the formation process of helical OPV-4 aggregates. Optical probing is a powerful technique on account of the extreme sensitivity of the π-conjugated chromophore to conformational, orientational, and supramolecular states of the OPV compound. We have analyzed the data with a model similar to the Oosawa-Kasai model for helical assembly of proteins in solution (*18*–*21*), which yields thermodynamic information such as the mean size of the assemblies as well as the equilibrium constants. Moreover, we uncovered the explicit role of the solvent molecules in the formation of the supramolecular stacks. Our findings can be generally applied to many chemical self-assembly processes of ordered quasi-one-dimensional stacks.

By using CD spectroscopy at a specific wavelength, we could monitor the optical changes characterizing the crossover from the monomeric state of OPV-4 to the aggregated state in dodecane. A Peltier-temperature programmer for thermostatting the samples was used to cool the solution slowly in small steps (from 375 K to 275 K, with a rate of 0.5 K/min and a resolution of one data point per 0.1 K). The slow cooling is necessary to suppress kinetic (i.e., nonequilibrium) effects on the self-assembly (fig. S2, A and B). We tracked the growth in the intensity of the λ = 466 nm CD band for four different OPV-4 concentrations and consistently observed a sharp increase at a specific temperature (fig. S2C), indicative of the transition from monomers into helical aggregates with a preferred handedness. This curve is not sigmoidal, and hence it cannot be described by an isodesmic model. However, it does follow expected Oosawa-Kasai behavior

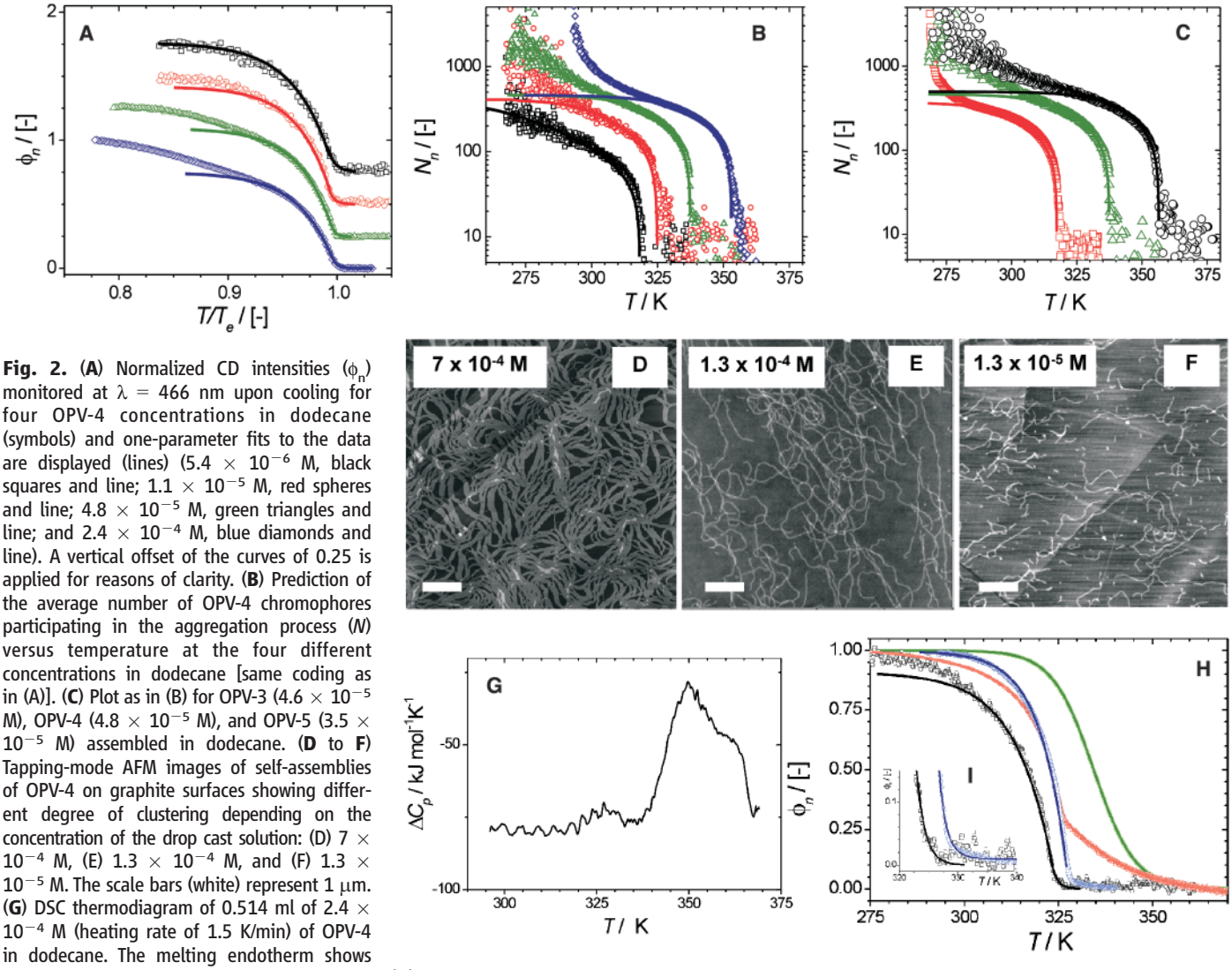


Fig. 2. (**A**) Normalized CD intensities (ϕ_n) monitored at λ = 466 nm upon cooling for four OPV-4 concentrations in dodecane (symbols) and one-parameter fits to the data are displayed (lines) (5.4 × 10^{-6} M, black squares and line; 1.1 × 10^{-5} M, red spheres and line; 4.8 × 10^{-5} M, green triangles and line; and 2.4 × 10^{-4} M, blue diamonds and line). A vertical offset of the curves of 0.25 is applied for reasons of clarity. (**B**) Prediction of the average number of OPV-4 chromophores participating in the aggregation process (*N*) versus temperature at the four different concentrations in dodecane [same coding as in (A)]. (**C**) Plot as in (B) for OPV-3 (4.6 × 10^{-5} M), OPV-4 (4.8 × 10^{-5} M), and OPV-5 (3.5 × 10^{-5} M) assembled in dodecane. (**D** to **F**) Tapping-mode AFM images of self-assemblies of OPV-4 on graphite surfaces showing different degree of clustering depending on the concentration of the drop cast solution: (D) 7 × 10^{-4} M, (E) 1.3 × 10^{-4} M, and (F) 1.3 × 10^{-5} M. The scale bars (white) represent 1 μm. (**G**) DSC thermodiagram of 0.514 ml of 2.4 × 10^{-4} M (heating rate of 1.5 K/min) of OPV-4 in dodecane. The melting endotherm shows transition temperatures at 351.0 K and 361.7 K. (**H**) Transition curves based on UV/vis [λ = 490 nm (green) and λ = 335 nm (blue)], fluorescence (λ = 500 nm), and CD (λ = 466 nm) spectra for OPV-4 (1.1 × 10^{-5} M). Fits to the data are shown on the basis of isodesmic or cooperative self-assembly. (**I**) A closer view of the nucleation regime based on the UV-vis (λ = 335 nm) and CD data.

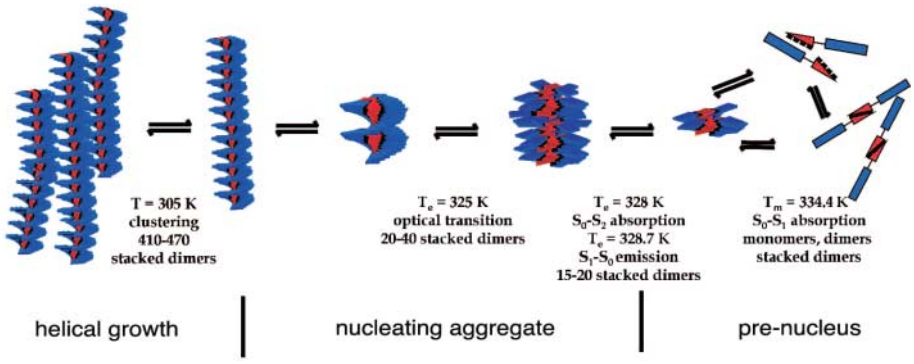


Fig. 3. Schematic representation of the hierarchical self-assembly of OPV-4 in solution based on the measurements presented in Fig. 2. First, monomers form dimers via quadruple hydrogen bonding. Upon cooling, about 10 to 15 dimers are brought together via an isodesmic pathway, forming disordered stacks. Upon further cooling, the molecules in the preaggregates become more restricted in relative position via a cooperative process (T_e = 328 K). In the next step, the preaggregates undergo a coil-helix transition to form a chiral nucleus of about 28 dimers, at which point the elongation-growth pathway sets in. Lastly, the cooperative stack length is reached and clustering of the assemblies occurs.

for thermally activated equilibrium polymerization (*19*), in which the non-isodesmic helical assembly (characterized by an activation step and subsequent propagation steps) is preceded by the isodesmic assembly of nonhelical structures. The modified model is a mathematically more tractable version of a self-assembled Ising chain theory introduced to describe the helix-coil transition in supramolecular polymers (*22*, *23*). In the simplified model, the two kinds of assemblies are linked through equilibrium between nonhelical and helical assemblies of a critical size (*19*). According to the model, the number average aggregation number of the helical assemblies N obeys the following relationship:

$$K_e = \exp[h_e(T - T_e)/RT_e^2] = 1 - N^{-1} + K_a N(N - 1) \quad (1)$$

with K_e as the equilibrium constant of the elongation process, K_a the equilibrium constant of the activation step, T the absolute temperature, h_e the molecular enthalpy of the elongation process, and R the gas constant. The fraction of molecules in assemblies, ϕ_n, obeys approximately

$$\{1 - \exp[h_e(T - T_e)/RT_e^2]\} \quad (2)$$

in the helical polymerized regime $T < T_e$ and approximates

$$\{K_a^{1/3}\exp[(\tfrac{2}{3}K_a^{-1/3} - 1)h_e(T - T_e)/RT_e^2]\} \quad (3)$$

near the polymerization temperature T_e [Supporting Online Material (SOM) Text]. Applying this model to our data, we can estimate N, the mean number of dimers in the aggregate, from Eq. 1; T_e, the temperature at which elongation sets in, ϕ_n, the normalized fraction of molecules in the aggregate, and h_e, the enthalpy of bond formation from both Eqs. 2 and 3; and K_a, the equilibrium constant between the active and nonactive state, from Eq. 3; active meaning the state that can initiate the elongation.

The normalized CD curves (Fig. 2A) are well fit by Eqs. 2 and 3, showing unambiguously that our model description is internally consistent and that helical aggregates are not present above the characteristic T_e (i.e., for values of T/T_e in excess of unity) but form only after the sharp nucleation step. The T_e decreases upon diluting the sample, revealing a linear relationship in the Van't Hoff plot (fig. S2D), with ΔH = –100 kJ/mol and ΔS = –215 J/mol K. Although there is a good correlation between experiment and theory in the case of the most dilute solution (5.4 μM) (Fig. 2A), at higher concentrations the theory underestimates the CD data at low temperatures (i.e., low T/T_e) (Fig. 2A). This result suggests an additional process that is not considered in the current polymerization theory. Interestingly, the fits for the different concentrations show that the temperature at which theory deviates from experiment (T_{sat}) rises with increasing concentration. This deviation is also visible in a plot of N versus temperature (Fig. 2B). At lower temperatures, the average columnar length estimated from the fitted model is remarkably constant for N values from 410 to 470. Assuming a π-π stacking distance of 0.35 nm (*15*), this N range corresponds to a length of 145 to 165 nm. These values are comparable with the values found by SANS and AFM [150 and 125 nm, respectively (table S1)]. With these data in hand, the deviation from one-dimensional growth at higher concentrations can be attributed as the clustering of stacks (both laterally and along the growing direction, Fig. 2, D to F). The data suggest this clustering is isodesmic in nature (fig. S3).

At the helical polymerization temperature $T = T_e$, the theory allows for the calculation of K_a: We find values of 2.0×10^{-4}, 0.46×10^{-4}, 0.31×10^{-4}, and 0.15×10^{-4}, in order from lowest to highest starting OPV concentra-

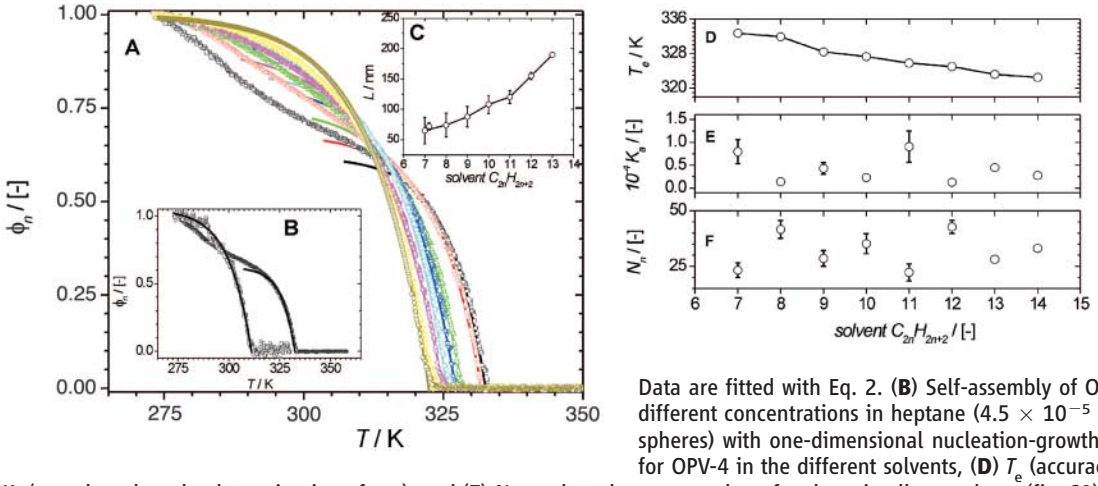


Fig. 4. Solvent-dependent CD measurements of OPV-4 in eight alkane solvents. (**A**) The normalized aggregate fraction of self-assembled OPV-4 in alkane solvents versus T/T_e [4.5×10^{-5} M $C_{2n}H_{2n+2}$, where n values were 7 (black), 8 (red), 9 (green), 10 (blue), 11 (cyan), 12 (magenta), 13 (yellow), and 14 (dark yellow)] based on monitoring CD intensity at λ = 466 nm upon cooling. Data are fitted with Eq. 2. (**B**) Self-assembly of OPV-4 versus temperature at two different concentrations in heptane (4.5×10^{-5} M, squares, and 4.5×10^{-6} M, spheres) with one-dimensional nucleation-growth fits. (**C**) Cooperative length (L) for OPV-4 in the different solvents, (**D**) T_e (accuracy is 0.1 K based on the fit), (**E**) K_a (error based on the determination of ϕ_{sat}), and (**F**) N are plotted versus number of carbons in alkane solvent (fig. S9). Error bars are based on the fit.

tion (Fig. 2A). The transition from the nonhelical state to the helical preaggregates is a prerequisite for spontaneous growth of the helix at temperatures below T_e (below $T/T_e = 1$) (fig. S2C); this behavior underlies the cooperative effect. From the plot in fig. S2E it is possible to estimate the average size of preaggregates or nuclei at $T = T_e$. The size of these nuclei, given by $N \sim K_a^{-1/3}$, thus yields chiral nuclei of 17 to 41 hydrogen-bonded dimers in order of ascending concentration at T_e. Computer simulations of the photophysical properties of the stacks suggested rotation angles between adjacent stacked dimers in the range of 6° to 12° (*24*). Intriguingly, in order to complete one helical pitch (180°), 15 to 30 stacked hydrogen-bonded dimers are needed. The rapid growth from these helical nuclei is caused by the many reinforcing noncovalent interactions and is enthalpy-driven, with an excess bond formation enthalpy of h_e = –56 kJ/mol estimated from the slope at T_e. The stability of the preaggregates is expected to increase when additional π-π interactions are provided by extending the chromophore length. Consistent with this reasoning, we observed that the T_e shifts from 317.5 K for OPV-3 to 325.0 K for OPV-4 and then to 356.5 K for OPV-5 (Fig. 2C) for comparable concentrations. Also, the length of the different molecular stacks can be estimated. The values of circa (ca.) 125 nm for the OPV-3 stack and ca. 175 nm for the OPV-5 stack bracket that of the OPV-4 stack (150 nm) and are in agreement with SANS data (table S1) (*15*).

The strong endotherm measured with ultrasensitive differential scanning calorimetry (DSC) confirms the high cooperativity of the polymerization. The DSC curve (Fig. 2G) shows the presence of two transitions, one broad transition from 340 to 365 K, coinciding with the absorption data (see below), and one sharp transition at 351 K. The latter coincides with the T_e extracted from the CD measurements. At these transition temperatures, we also observe the largest volume changes in preliminary pressure perturbation calorimetric experiments. Further research is needed to correlate the thermodynamic parameters, acquired with the different techniques, to the individual steps in the self-assembly process (Fig. 3). To obtain a more detailed picture of the characteristics of the nuclei, we turned to fluorescence and UV/vis spectroscopy, again using the same slow cooling rate of 0.5 K/min to suppress kinetic (nonequilibrium) effects (Fig. 2H). Careful analysis of the UV/vis and fluorescence data (Fig. 2, H and I, and SOM Text) reveals experimental evidence for both isodesmic (by using eq. S3) and cooperative transitions (by using Eqs. 2 and 3) at different temperatures (figs. S4 to S6).

Combining all data, it is possible to postulate a hierarchical pathway for the self-assembly of OPV-4 into fibrillar structures upon cooling a solution of molecularly dissolved monomers at high temperature (Fig. 3). First, the monomers form dimers via quadruple hydrogen bonding. The association of the very first hydrogen-bonded dimers into short stacks having disorder both inside the molecules as well as in the stacking direction follows an isodesmic pathway. This isodesmic equilibrium state continues to shift upon cooling until 10 to 15 stacked dimers are brought together, characterized by sigmoidal transitions in the UV/vis spectrum [T_m = 334.4 K, absorption at λ = 490 nm, $S_0 \rightarrow S_1$ transition (Fig. 2H)] and fluorescence data (first part of the melting curve in Fig. 2H). Upon further cooling, the molecules in the preaggregates become more restricted in their relative positions, characterized by the abrupt change at T_e = 328 K in the absorption at λ = 335 nm [more localized $S_0 \rightarrow S_2$ transition (*25*) (Fig. 2, H and I)] and in the fluorescence spectra [(Fig. 2H) quenching is presumably related to faster diffusion of excitons to trap sites (*26*) due to improved ordering (fig. S6)]. This cooperative transition is close to the onset of chiroptical activity in the preaggregates (Fig. 2I). Helix formation transforms the preaggregate into a chiral nucleus of 28 dimers (fig. S4) (at a concentration of 1.1×10^{-5} M, $K_a = 0.46 \times 10^{-4}$) before the elongation-growth pathway sets in, as detected by the cooperative CD transition. The persistence length of the structures is independent of concentration but depends on the size of the molecule. Concentration, on the other hand, determines the temperature at which the onset of clustering occurs.

We have found that solvent structure dramatically influences the stability of the nuclei and stacks (Fig. 4). Upon changing the length of the alkane solvent in the self-assembly process (fig. S7), the T_e shifts from 322.5 (tetradecane) to 332.7 K (heptane) as measured at constant chromophore concentration (Fig. 4A). The observed change is not linear but shows an oscillatory dependence on whether the solvent contains an odd or even number of carbons (Fig. 4D). When K_a as well as N is plotted versus the number of C atoms in the solvent, a strong odd-even effect again is observed (Fig. 4, E and F). Nearly all physical parameters in alkane solvents depend linearly on the number of carbons in the alkane chain except for properties related to ordering, such as the density and melting points of the alkanes. It therefore appears that coorganization of the solvent at the periphery of the aggregates plays a direct role in the assembly process, evident even during the formation of the preaggregates. Whereas the full assembly curve measured in tetradecane can be fitted with a one-dimensional growth model, this model underestimates the experimental data measured in other solvents. The point of deviation, however, occurs at different temperatures depending on the length of the solvent molecule. In the shortest solvent chains, the self-assembly is apparently by far the most sensitive for clustering, and the average columnar length is shortest (Fig. 4C and fig. S8). This finding has important consequences for understanding the different properties found for the same molecular stacks formed in different solvents and suggests that the solvent does not play a passive role in supramolecular assembly, but rather must be considered explicitly.

On one hand, our study shows that the expected nucleation-growth mechanism is applicable to chemical self-assembly structures. On the other hand, unexpected details about the nucleating species and persistency and the role of the solvent have become visible. All of these findings are potentially relevant to the formation processes of fibers, gels, and other supramolecular structures in general.

References and Notes

1. G. M. Whitesides, M. Boncheva, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4769 (2002).
2. P. Terech, R. G. Weiss, *Chem. Rev.* **97**, 3133 (1997).
3. T. Shimizu, M. Masuda, H. Minamikawa, *Chem. Rev.* **105**, 1401 (2005).
4. F. J. M. Hoeben, P. Jonkheijm, E. W. Meijer, A. P. H. J. Schenning, *Chem. Rev.* **105**, 1491 (2005).
5. J. J. L. M. Cornelissen, A. E. Rowan, R. J. M. Nolte, N. A. J. M. Sommerdijk, *Chem. Rev.* **101**, 4039 (2001).
6. A. Lohr, M. Lysetska, F. Würthner, *Angew. Chem. Int. Ed. Engl.* **44**, 5071 (2005).
7. R. F. Service, *Science* **309**, 95 (2005).
8. B. Moulton, M. J. Zaworotko, *Chem. Rev.* **101**, 1629 (2001).
9. P. J. Flory, *J. Chem. Phys.* **17**, 223 (1949).
10. A. Kentsis, K. L. B. Borden, *Curr. Protein Peptide Sci.* **5**, 125 (2004).
11. A. Klug, *Philos. Trans. R. Soc. London Ser. B* **354**, 531 (1999).
12. D. L. Caspar, *Biophys. J.* **32**, 103 (1980).
13. R. Krishnan, S. L. Lindquist, *Nature* **435**, 765 (2005).
14. A. P. H. J. Schenning, P. Jonkheijm, E. Peeters, E. W. Meijer, *J. Am. Chem. Soc.* **123**, 409 (2001).
15. P. Jonkheijm *et al.*, *J. Am. Chem. Soc.* **125**, 15941 (2003).
16. A. Gesquière *et al.*, *Nano Lett.* **4**, 1175 (2004).
17. C. R. L. P. N. Jeukens *et al.*, *J. Am. Chem. Soc.* **127**, 8280 (2005).
18. F. Oosawa, M. Kasai, *J. Mol. Biol.* **4**, 10 (1962).
19. P. van der Schoot, in *Supramolecular Polymers*, A. Ciferri, Ed. (CRC Press, Baton Rouge, LA, 2005).
20. A. Aggeli *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 11857 (2001).
21. F. Oosawa, S. Asakura, *Thermodynamics of the Polymerization of Proteins* (Academic Press, London, 1975).
22. J. van Gestel, P. van der Schoot, M. A. J. Michels, *Langmuir* **19**, 1375 (2003).
23. P. van der Schoot, M. A. J. Michels, L. Brunsveld, R. P. Sijbesma, A. Ramzi, *Langmuir* **16**, 10076 (2000).
24. D. Beljonne *et al.*, *J. Phys. Chem. B* **109**, 10594 (2005).
25. L. Herz *et al.*, *Phys. Rev. B* **68**, 045203 (2003).
26. E. Peeters, A. Marcos Ramos, S. C. J. Meskers, R. A. J. Janssen, *J. Chem. Phys.* **112**, 9445 (2000).

27. P.J. was a postdoctoral fellow in the framework of BIOMADE (EU). This work was further financially supported by grants from the Dutch Council for Sciences (NWO). We thank F. J. M. Hoeben for providing the OPV-3; R. A. J. Janssen, L. Brunsveld, M. van Genderen, and M. Smulders for stimulating discussions; and L. Mitra and R. Winter for collaborating and measuring preliminary calorimetric measurements.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/80/DC1
Materials and Methods
SOM Text
Figs. S1 to S9
Table S1
References

24 March 2006; accepted 22 May 2006
10.1126/science.1127884

Photoproduction of Proton Gradients with π-Stacked Fluorophore Scaffolds in Lipid Bilayers

Sheshanath Bhosale,[1] Adam L. Sisson,[1] Pinaki Talukdar,[1] Alexandre Fürstenberg,[2] Natalie Banerji,[2] Eric Vauthey,[2] Guillaume Bollot,[1] Jiri Mareda,[1] Cornelia Röger,[3] Frank Würthner,[3] Naomi Sakai,[1] Stefan Matile[1]*

Rigid *p*-octiphenyl rods were used to create helical tetrameric π-stacks of blue, red-fluorescent naphthalene diimides that can span lipid bilayer membranes. In lipid vesicles containing quinone as electron acceptors and surrounded by ethylenediaminetetraacetic acid as hole acceptors, transmembrane proton gradients arose through quinone reduction upon excitation with visible light. Quantitative ultrafast and relatively long-lived charge separation was confirmed as the origin of photosynthetic activity by femtosecond fluorescence and transient absorption spectroscopy. Supramolecular self-organization was essential in that photoactivity was lost upon rod shortening (from *p*-octiphenyl to biphenyl) and chromophore expansion (from naphthalene diimide to perylene diimide). Ligand intercalation transformed the photoactive scaffolds into ion channels.

In biological systems, light harvesting uses the energy in photons to drive chemical reactions. Although many biomimetic and supramolecular compounds have been devised to collect photons, the output of these systems often consists of photons of lower energy. The outstanding challenge is to devise schemes in which acceptor chromophores can transfer the charge carrier (electron or hole) to generate a chemical product (for example, H^+ or O_2) before relaxation to the initial state occurs.

In proteins, a series of chromophores is used to separate charges and avoid back reactions (*1*–*3*). In smaller biomimetic assemblies, an alternative strategy is to transfer charge to electron donors and acceptors in solution, provided that these donors and acceptors can be spatially separated and that these transfers occur faster than the charge recombination (*4*–*7*). We report the design, synthesis, and evaluation of blue, red-fluorescent rigid-rod photosystems (**1**), in which the π-stack apparently supports sufficient charge-separation lifetimes for conversion into chemical gradients (Fig. 1).

The construction of complex and multifunctional π-architecture is difficult because of the poor directionality of π,π-interactions (*8*–*18*). Apart from a central role in oligonucleotide chemistry and biology (*18*), cofacial π-stack architecture contributes very little to electron transfer in biological (*1*–*3*) and synthetic (*4*–*7*) photosystems, not to speak of the structure of biological or synthetic ion channels or pores (*19*). Recently, we have demonstrated that the rational design of refined transmembrane π-stack architecture is feasible with the use of rigid-rod molecules as preorganizing scaffolds (*20*, *21*). This strategy was applied not only to create photosystem **1** but also to introduce multifunctionality; that is, to enable photosystem **1** to open up into ion channel **2** in response to the intercalation of ligands **3** (Fig. 1).

Photosystem **1** was prepared by self-assembly of four *p*-octiphenyls (**4**), each bearing eight core-substituted naphthalene diimides (NDIs) (*22*) along the rigid-rod scaffold (Fig. 2 and scheme S1). The NDI octamer **4** was synthesized from simple commercial starting materials **5** to **10** in 20 steps altogether. Pyrene **5** was oxidized to the core-chlorinated dianhydride to allow for the introduction of terminal (**6** and **7**) and central (**8**) amines before attachment along the *p*-octiphenyl rod made from four biphenyls (**10**) (*23*). The following control molecules were synthesized in a similar manner: NDI dimer **11**, NDI monomer **12**, perylene diimide (PDI) octamer **13**, and PDI dimer **14** (schemes S2 and S3).

Unsubstituted NDIs are colorless nonfluorescent electron traps. However, recent findings show that a minor change on the structural level suffices to introduce all of the characteristics needed for photosynthetic activity. Core substitution with electron-donating alkylamines produces compact blue, red-emitting push-pull fluorophores (excitation wavelength λ_{ex} = 638 nm; emission wavelength λ_{em} = 678 nm), lowers the reduction potential by 300 mV, and moves the oxidation potential within reach (*22*). The resulting potentials for NDI reduction to the radical anion $NDI^{\bullet-}$ [redox potential ($E_{1/2}$) = –710 mV] and oxidation to $NDI^{\bullet+}$ ($E_{1/2}$ = +1290 mV) appeared compatible with photoinduced charge separation in π-stack architecture **1** for electron transfer to

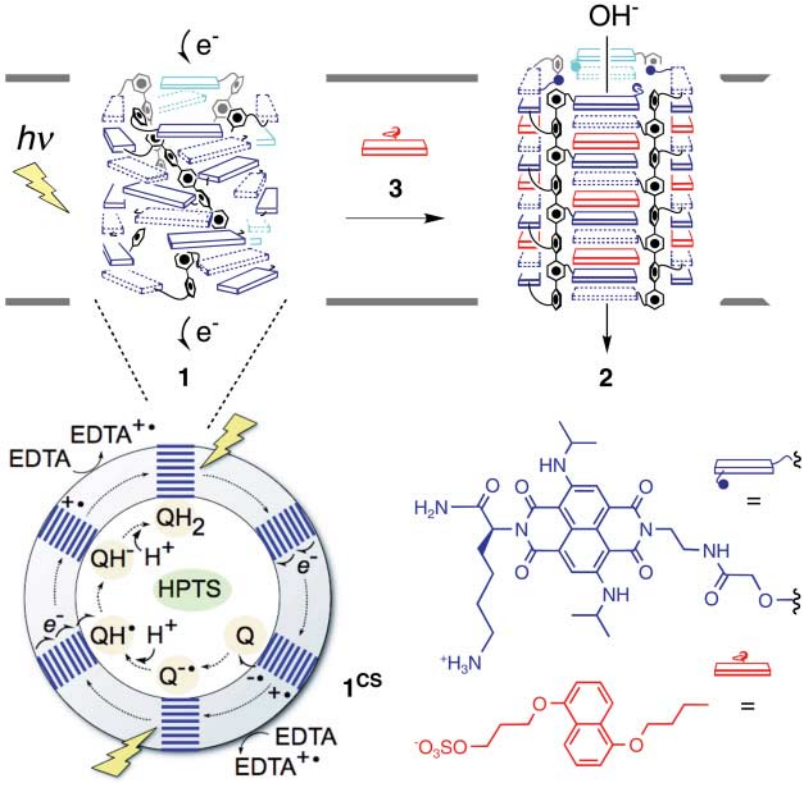


Fig. 1. Smart photosystem **1** with an indication of the ligand-gated opening into ion channel **2** and fluorometric detection of photosynthetic activity in bilayer membranes with HPTS as internal increase in pH after photoinduced charge separation ($\mathbf{1}^{CS}$), followed by external EDTA oxidation and internal reduction of quinones (Q). *hν*, photonic energy.

[1]Department of Organic Chemistry, [2]Department of Physical Chemistry, University of Geneva, Geneva, Switzerland. [3]Institut für Organische Chemie, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany.

*To whom correspondence should be addressed. E-mail: stefan.matile@chiorg.unige.ch

2-sulfonato-1,4-naphthoquinone (Q) ($E_{1/2}$ = –60 mV) and hole transfer to EDTA ($E_{1/2} \approx$ +430 mV) (table S1). This photoredox process is thermodynamically unfavorable and thus converts photonic energy into chemical energy.

The photosynthetic activity of quadruple π-*M*-helices **1** was measured in egg yolk phosphatidylcholine large unilamellar vesicles (EYPC LUVs). The vesicles were loaded with quinone acceptor Q and pH detector 8-hydroxypyrene-1,3,6-trisulfonate (HPTS) and surrounded with the electron donor EDTA (Fig. 1) (*4*, *20*). Internal quinone reduction in response to the irradiation of the externally added NDI rods **4** at 635 nm was detectable as an increase in intravesicular pH ("a" in Fig. 3A and fig. S4). The initial velocity of change in HPTS emission, which reflects internal proton consumption, was analyzed with Hill's equation. The value of n = 3.9 ± 1.2 that was found indicated that the active photosystem **1** is a tetrameric supramolecule with a median effective concentration of 1.3 μM (Fig. 3B) (*24*). Judged from exciton coupling between twisted proximal NDI chromophores throughout circular dichroism (CD) spectra (*21*), this tetramer is a heat-insensitive (melting temperature ≥ 70°C) *M*-helix with sensitivity toward chemical denaturation [with guanidinium chloride, free energy of self-organization (ΔG^{H2O}) ≤ 1.17 kcal/mol, fig. S7]. Support for cofacial π-stacking (*8*) in quadruple π-*M*-helix **1** was secured with the appearance of a hypsochromic maximum at 598 nm that was absent in the absorption spectrum of monomer **12** and was much weaker with dimer **11** (fig. S9).

Time-resolved fluorescence and transient absorption measurements provided direct experimental evidence that, upon irradiation of **4**, a charge-separated state characterized by a relatively long lifetime (61 ps with **4** versus 23 ps with **11**) is populated almost quantitatively (>97%, compared to 68% with **11**) on an ultrafast time scale (with a distribution of time constants between <2 and ~7 ps) (Fig. 4 and figs. S12 to S17). These features account for the marked activity of photosystem **1**. The appearance of new bands in the transient absorption spectrum of the NDI monomer **12** (trace a in Fig. 4) in the presence of the electron donor *N*,*N*-dimethylaniline (DMA) confirmed that the radical anion $NDI^{\bullet-}$ was being detected (trace b in Fig. 4). The presence of the $NDI^{\bullet-}$ bands in the transient absorption spectra of dimer **11**, as well as bands of the more pronounced, more stable octamer **4**, supported photoinduced population of charge-separated state $\mathbf{1}^{CS}$ (trace c in Fig. 4).

The addition of intercalator **3** caused a ligand-gated helix-barrel transition from photosystem **1** to ion channel **2** (fig. S5) (*20*, *21*) and thus resulted in the deletion of light-induced proton gradients [Figs. 1 and 3A (parts a and b)]; addition of the ion channel gramicidin A caused the same result. Photosynthetic activity was naturally inhibited by intercalator **3** (part d in Fig. 3A). The helix-barrel transition by intercalator **3** was also reflected in the expected CD silencing (*21*, *25*), together with a small increase in emission intensity that may be interpreted as hindered charge separation in barrel-stave complex **2** (figs. S6 and S8).

Control molecules **11**, **13**, and **14** were made to elucidate the dependence of photosynthetic activity of photosystem **1** on rod contraction from *p*-octiphenyl to biphenyl and on chromophore expansion from NDI to PDI. Inactivity of NDI dimer **11** (part c in Fig. 3A) coincided with CD silencing, the partial disappearance of the 598-nm absorption indicative of cofacial π-stacking, and incomplete photoinduced charge separation with a reduced lifetime (Fig. 4 and figs. S9 to S17). Notably, in contrast, the independence of both chiroptical and poor photosynthetic activity on rod length was found for the PDI series represented by octamer **13** and dimer **14** (fig. S20 and table S4). In agreement with molecular models (Fig. 5 and figs. S18 and S19), these results confirmed that the smart rigid-rod π-stack NDI nanoarchitecture of photosystem **1** is distinctive (that is, incompatible with the PDI series) and essential for function.

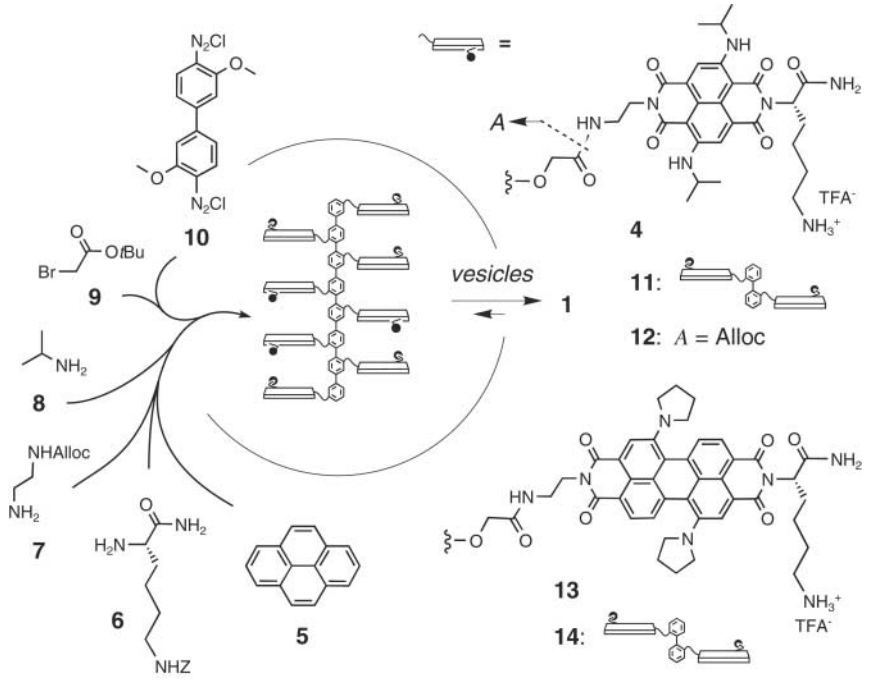


Fig. 2. Synthesis of rod **4** for self-assembly into photosystem **1** in lipid bilayers (Fig. 1) and structure of control molecules **11** to **14**.

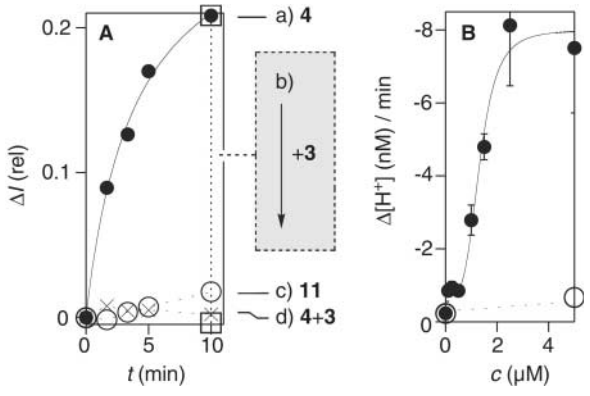


Fig. 3. Photoactivity of NDI octamer **4** (a, solid circles) with in situ opening into ion channel **2** by addition of ligand **3** (b, open squares), and comparison to NDI dimer **11** (c, open circles) and preformed ion channel **2** (**4**+**3**, d,×). (**A**) Change in HPTS excitation ratio I_{462}/I_{405} (λ_{em} = 510 nm) as a function of irradiation time t (λ_{ex} = 635 nm) in the presence of **11** (10 μM) or **4** (1.5 μM) in EYPC LUVs (~65 μM) with external EDTA (5 mM; 0.95 mM $K_nH_{3-n}PO_4$, 95 mM KCl, pH 7.0) and internal quinone (10 mM) and HPTS (1 mM; 100 mM KCl, pH 7.0); ligand **3** (20 μM, or gramicidin A) was added at t = 10 min or t = 0 min. (**B**) Initial velocity of proton consumption as a function of the concentration c of **4** (solid circles) and **11** (open circles). Error bars indicate SE.

References and Notes

1. J. Deisenhofer, H. Michel, *Science* **245**, 1463 (1989).
2. P. Jordan *et al.*, *Nature* **411**, 909 (2001).
3. N. Nelson, A. Ben-Shem, *Nat. Rev. Mol. Cell Biol.* **5**, 971 (2004).

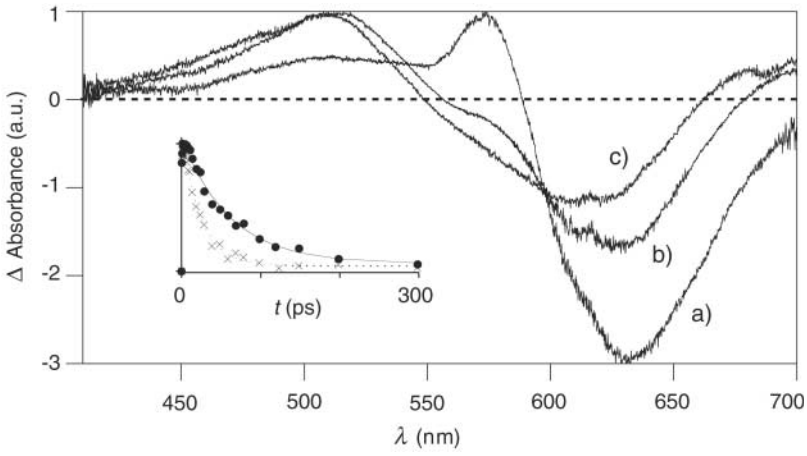


Fig. 4. Intensity-normalized transient absorption spectra of NDI octamer **4** (trace c) and NDI monomer **12** in the presence (trace b) or absence (trace a) of DMA in methanol after excitation with a 50-fs laser pulse at 600 nm. Δ absorbance, change in absorbance; a.u., arbitrary units. The bands at 510 and above 670 nm seen with **4** and **12** in the presence of DMA are attributed to the $NDI^{\bullet-}$ radical, whereas with **12** in the absence of DMA, the 450- to 580-nm band originates from singlet excited-state absorption. The negative signal between 600 and 670 nm is due to ground-state bleaching and stimulated emission. (**Inset**) Time profiles of the absorption due to the radical anion of octamer **4** (circles) and dimer **11** (×). The lines represent best fits to the data.

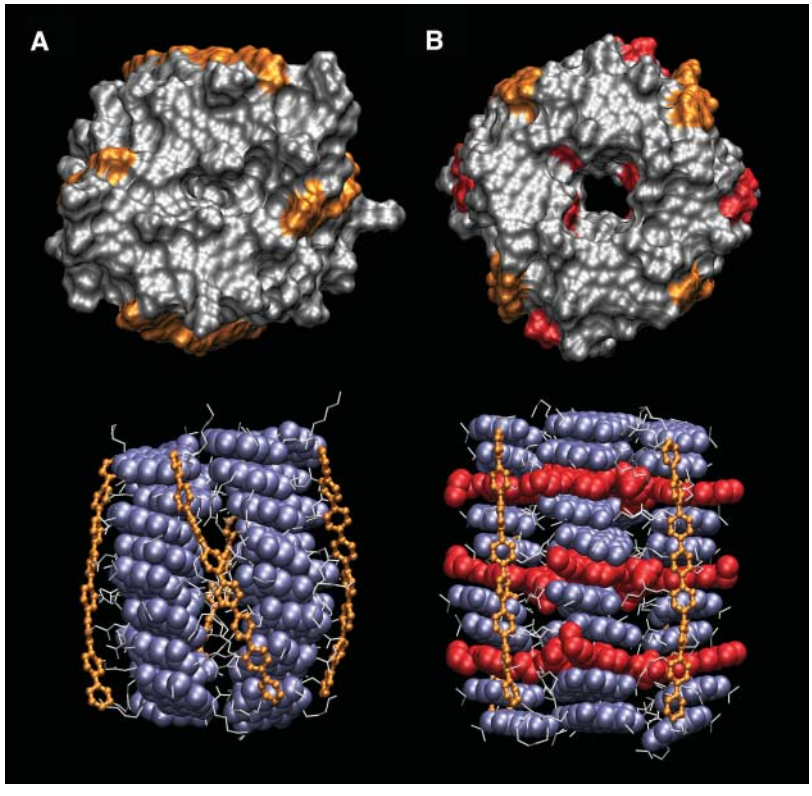


Fig. 5. Molecular dynamics simulations of photosystem **1** (**A**) and ion channel **2** (**B**) with 12 ligands **3** (red) in axial view (top) and side view (bottom). In (A) (closed pore), helicity (M) = –50.2° and height (h) = 32 Å; in (B), inner diameter = 4.8 Å, M = –8.8°, and $h \approx 40$ Å.

4. G. Steinberg-Yfrach *et al.*, *Nature* **385**, 239 (1997).
5. J. J. Grimaldi, S. Boileau, J.-M. Lehn, *Nature* **265**, 229 (1977).
6. J. N. Robinson, D. J. Cole-Hamilton, *Chem. Soc. Rev.* **20**, 49 (1991).
7. L. Zhu, R. F. Khairutdinov, J. L. Cape, J. K. Hurst, *J. Am. Chem. Soc.* **128**, 825 (2006).
8. J. M. Giaimo, A. V. Gusev, M. R. Wasielewski, *J. Am. Chem. Soc.* **124**, 8530 (2002).
9. M. J. Ahrens *et al.*, *J. Am. Chem. Soc.* **126**, 8284 (2004).
10. L. L. Miller, K. R. Mann, *Acc. Chem. Res.* **29**, 417 (1996).
11. H. E. Katz *et al.*, *Nature* **404**, 478 (2000).
12. J. M. Warman *et al.*, *Chem. Mater.* **16**, 4600 (2004).
13. F. J. M. Hoeben, P. Jonkheijm, E. W. Meijer, A. P. H. J. Schenning, *Chem. Rev.* **105**, 1491 (2005).
14. V. Balzani, M. Venturi, A. Credi, *Molecular Devices and Machines* (Wiley, Weinheim, Germany, 2003), chap. 6.
15. L. Schmidt-Mende *et al.*, *Science* **293**, 1119 (2001).
16. R. S. Lokey, B. L. Iverson, *Nature* **375**, 303 (1995).
17. F. Würthner, Ed., *Supramolecular Dye Chemistry* (Topics in Current Chemistry Series **258**, Springer, Berlin, 2005).
18. C. Wan, T. Fiebig, O. Schiemann, J. K. Barton, A. H. Zewail, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 14052 (2000).
19. S. Matile, A. Som, N. Sordé, *Tetrahedron* **60**, 6405 (2004).
20. P. Talukdar, G. Bollot, J. Mareda, N. Sakai, S. Matile, *J. Am. Chem. Soc.* **127**, 6528 (2005).
21. P. Talukdar, G. Bollot, J. Mareda, N. Sakai, S. Matile, *Chem. Eur. J.* **11**, 6525 (2005).
22. F. Würthner, S. Ahmed, C. Thalacker, T. Debaerdemaeker, *Chem. Eur. J.* **8**, 4742 (2002).
23. See supporting material on *Science* Online.
24. To confirm the validity of this result on the structural level, a CD method for the determination of meaningful Hill coefficients for exergonic self-assembly under denaturating conditions was introduced (*26*).
25. J. M. Kerckhoffs *et al.*, *J. Am. Chem. Soc.* **127**, 12697 (2005).
26. S. Bhosale, S. Matile, *Chirality*, in press.

27. We thank D. Jeannerat, A. Pinto, and S. Grass for nuclear magnetic resonance measurements; P. Perrottet and the group of F. Gülaçar for mass spectrometry; and the Swiss National Science Foundation for financial support (to S.M. and E.V.). C.R. thanks the Degussa Foundation for a Ph.D. scholarship, and S.B. is a fellow of the Roche Research Foundation.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/84/DC1
Materials and Methods
SOM Text
Figs. S1 to S20
Schemes S1 to S3
Tables S1 to S4
References

22 February 2006; accepted 16 May 2006
10.1126/science.1126524

Hydrogen Isotopes in Eocene River Gravels and Paleoelevation of the Sierra Nevada

Andreas Mulch,* Stephan A. Graham, C. Page Chamberlain

We determine paleoelevation of the Sierra Nevada, California, by tracking the effect of topography on precipitation, as recorded in hydrogen isotopes of kaolinite exposed in gold-bearing river deposits from the Eocene Yuba River. The data, compared with the modern isotopic composition of precipitation, show that about 40 to 50 million years ago the Sierra Nevada stood tall (≥2200 meters), a result in conflict with proposed young surface uplift by tectonic and climatic forcing but consistent with the Sierra Nevada representing the edge of a pre-Eocene continental plateau.

The evolution of mountainous topography reflects a balance between tectonic forces in Earth's lithosphere and upper mantle and erosion at Earth's surface. However, their relative roles remain controversial (*1*, *2*). Understanding the long-term topographic history of the Sierra Nevada, California, figures importantly in this debate. One view is that the bulk of elevation gain took place in the Pliocene about 3 to 5 million years ago (Ma) (*3*–*5*) and is the result of removal of dense mantle lithosphere (*6*–*8*). In contrast, the Sierra Nevada may have existed as a major topographic feature throughout the Cenozoic (*9*–*13*) and the apparent young uplift may be due, in part, to climatic changes at 3 to 5 Ma (*14*, *15*). To provide a further test, we determine Eocene (~40 to 50 Ma) stream gradients and headwater elevations by measuring the hydrogen isotope composition of kaolinite from Eocene fluvial sediments that record the isotopic composition of Eocene surface waters from ancient sea level continuously upstream into the modern (and ancient) range. One key piece of evidence cited for considerable (>1000 m) Pliocene surface uplift (*5*) comes from Eocene river gravels exposed along the western flank of the Sierra Nevada. These gravels, mined for gold in the later half of the 1800s, are nearly continuously exposed in open-cast mines from Eocene sea level upstream to current elevations of 2000 m, where they comprise the fill of bedrock-incised valleys and canyons, locally greater than 200 m deep. In this ancient fluvial system (*16*, *17*), kaolinitized Eocene river deposits are overlain by Oligocene (38.9 Ma) volcanoclastic units, with the kaolinite formed during Eocene weathering of granitic fluvial detritus shortly after deposition in the ancient Yuba River (*18*). Because kaolinite incorporates water into the crystal structure during its formation, its hydrogen isotopic composition (δD_{kaol}) serves as a proxy for the isotopic composition of Eocene precipitation and near-surface ground water along the river channel. As in orogenic belts, the isotopic composition of precipitation scales with altitude, δD_{kaol} values are thus largely controlled by changes in elevation (*19*–*22*).

We determined 44 δD_{kaol} values (table S1) from 15 localities along the Eocene Yuba and American rivers (Fig. 1). We collected in situ kaolinite from individual granitic clasts, as well as samples from Eocene channel sands, which reflect detrital kaolinite transported from upstream. Both are from the upper part of the depositional sequence (commonly referred to as bench gravels) (fig. S2) (*16*, *17*). The difference in δD_{kaol} from Eocene sea level to upstream weathering sites, $\Delta(\delta D_{kaol})$, provides an estimate of paleoelevation because the change in isotopic composition of precipitation, $\Delta(\delta D_{ppt})$, during ascent of cloud systems is related to the change in elevation (*19*, *20*, *22*). We convert $\Delta(\delta D_{kaol})$ to paleoelevation using a thermodynamic model (*20*) for isotopic lapse rates (fig. S1). This model includes Eocene sea surface temperature and relative humidity as inputs and reproduces equilibrium fractionation of $\delta^{18}O_{ppt}$ and δD_{ppt} between precipitation and water vapor, based on Rayleigh distillation of an air mass during orographic ascent (*18*). For elevations of 1000 to 2000 m, the modeled Eocene hydrogen isotope lapse rate (–1.65‰/100 m) and the measured modern lapse rate (–1.70‰/100 m) (*23*) are identical within error. The methodological error on the paleoelevation estimates is ~ +350/–400 m (fig. S1).

For all three tributaries of the Eocene Yuba and adjacent American rivers, δD of in situ kaolinite decreases systematically upstream (Fig. 2) from δD_{kaol} = –88‰ at 400 m of modern elevation to δD_{kaol} = –105‰ at

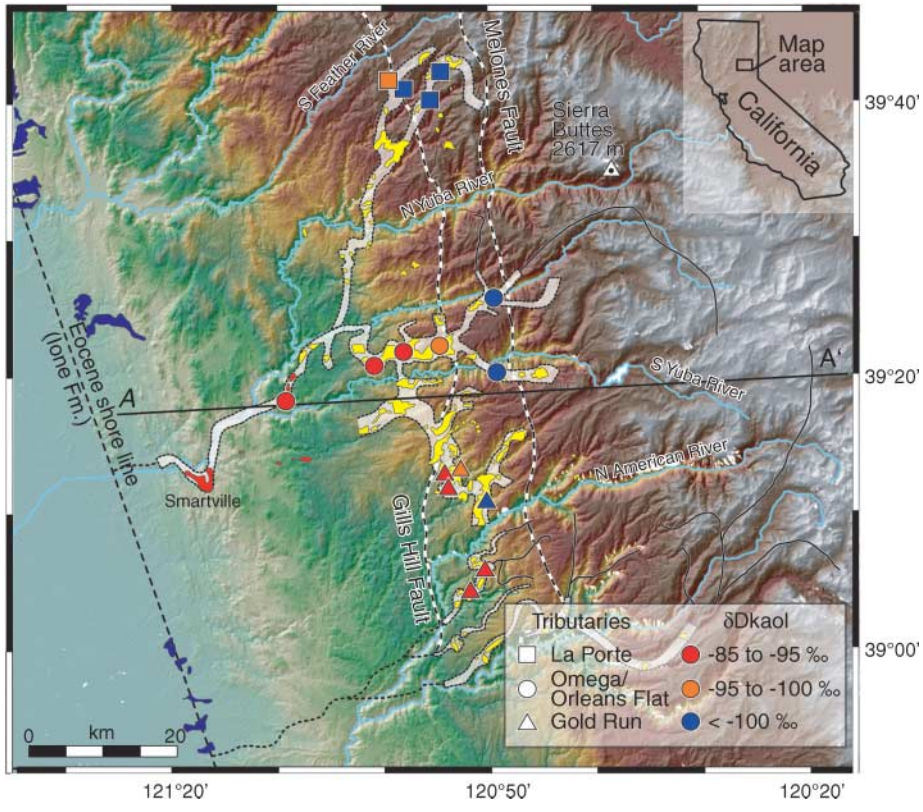


Fig. 1. Shaded relief map of northern Sierra Nevada. Sampling locations of kaolinitized river gravels shown by filled symbols and color-coded according to mean δD_{kaol} value. Eocene shoreline (*9*) is defined by coastal to marine Ione Formation and temporally equivalent to deposition of gravels in the ancestral Yuba River. Reconstructed course of Eocene Yuba River (white) based on exposure of Eocene gravel (yellow). A-A′ is profile in Fig. 3.

Geological and Environmental Sciences, Stanford University, 450 Serra Mall, Building 320, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: mulch@pangea.stanford.edu

~1600 m of modern elevation. In contrast, δD of detrital kaolinite is not correlated with distance from the Eocene shoreline but is identical to the most deuterium-depleted in situ kaolinite at the highest reaches of the Yuba River (Fig. 2). This result is expected, as detrital kaolinite should be transported downstream from its weathering location to the site of deposition. Systematic differences in δD of detrital and in situ kaolinite document that δD_{kaol} values reflect the conditions during Eocene weathering and that any postdepositional processes did not affect the δD_{kaol} in the river deposits (*18*).

Converting these isotopic compositions to paleoelevation requires knowledge of δD_{ppt} at a low elevation site. However, in situ kaolinite is not exposed in the lowest segments of the ancient Yuba River that we sampled, and the clay mineral inventory is dominated by detrital kaolinite. In the absence of isotopic data for kaolinite at Eocene sea level, we use a value based on a linear fit through δD_{kaol} (in situ) that gives a sea level intercept at $\delta D_{kaol} = -75 \pm 5‰$ ($r^2 = 0.76$) (Fig. 2). This value is lower than δD_{kaol} in equilibrium with precipitation recorded at most low-elevation, low-latitude (<35°) stations in the modern International Atomic Energy Agency (IAEA) Global Network of Isotopes in Precipitation (*18*, *24*). For the most conservative paleoelevation estimate we, therefore, adopt the lower bound on the sea level intercept (δD_{kaol} = –80‰) as our sea level δD_{kaol}, which results in lower estimated paleoelevations.

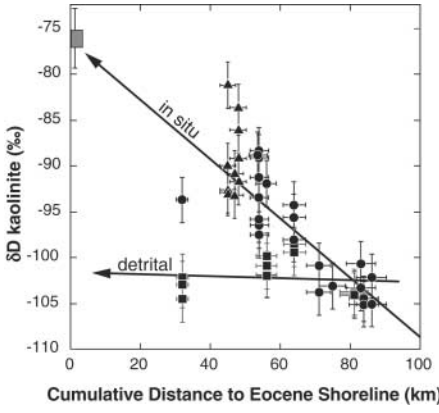


Fig. 2. δD_{kaol} along Eocene Yuba (circles) and American (triangles) rivers versus distance to Eocene shoreline. Linear regression through in situ δD_{kaol} from Yuba River indicates sea level $\delta D_{kaol} = -76 \pm 5‰$ ($r^2 = 0.76$) within error identical to kaolinite in isotopic equilibrium with modern precipitation at the Pacific coast (gray square). Detrital kaolinite (squares) overlaps with most D-depleted in situ kaolinite. The 25 ± 5‰ decrease in δD_{kaol} from sea level upstream reflects rainout during uplift of moisture along the western flank of the Eocene Sierra Nevada.

Based on the measured δD_{kaol} in the upstream regions of the La Porte, Omega, and Gold Run tributaries (Fig. 1), the average $\Delta(\delta D_{kaol})$ is –25‰ (Fig. 2). This value corresponds to an average Eocene elevation of the highest exposed ancestral Yuba River deposits of ~1.5 km (1503 +379/–498 m, including a 5‰ uncertainty on the δD_{kaol} at sea level) (Fig. 3) (*18*). Given that these deposits are found downstream of Eocene bedrock canyons, this elevation likely underestimates paleoelevation of the headwaters of the Eocene Yuba River, as low-δD waters must have flowed from adjacent mountaintops.

Three results come from our paleoaltimetry analysis. First, Eocene elevations for different tributaries of the ancestral Yuba River overlap within error with modern elevations. This result, seemingly at odds with paleoelevations inferred from studies based on differential tilt of Eocene and younger strata (*3*–*5*), suggests that the western slopes of the Eocene and the modern northern Sierra Nevada were similar.

Second, there is a larger scatter of calculated elevations observed at 40 to 60 km distance from the Eocene shoreline (Fig. 2). We suggest that this scatter reflects the hypsometry of the drainage patterns along the northwestward-trending aggradational segment of the ancient Yuba where the course of the river follows two major fault zones (Fig. 1). Within this northwest trending bedrock-controlled segment, we expect to see mixing of surface runoff from steep west-directed tributaries that drain high elevations to the east and meteoric water that reflects the local hypsometry along the more shallowly dipping range-parallel river segment.

Third, the result that reconstructed elevations for all three tributaries are identical suggests that they were draining a tall range with moderate relief. This is consistent with the measured Eocene bedrock incision that is commonly ~200 to 300 m but up to 790 m near the Sierra Buttes (*4*) (Fig. 1). Together with the 1503 +379/–498 m estimate for the elevation of the Eocene river bed, the bedrock incision data imply that Eocene elevations were 1.7 to 1.8 km in the northern Sierra Nevada and that peaks reached up to ~2200 m above sea level. Both the modern Yuba River headwaters (~2100 m) and peak elevations (~2780 m) in the northern Sierra Nevada are ~25% higher than in the Eocene, indicative of post-Eocene uplift of 350 to 600 m. This result is in agreement with estimates of 300 to 500 m of Pliocene (3 to 5 Ma) surface uplift of the Sierra Nevada as inferred from measured river incision rates (*14*). Because relative surface displacements between the western Basin and Range province and the northern Sierra Nevada are of Miocene and younger age (*25*), we speculate that the Eocene Sierra Nevada formed the western edge of a high-elevation landscape that characterized large areas of the western United

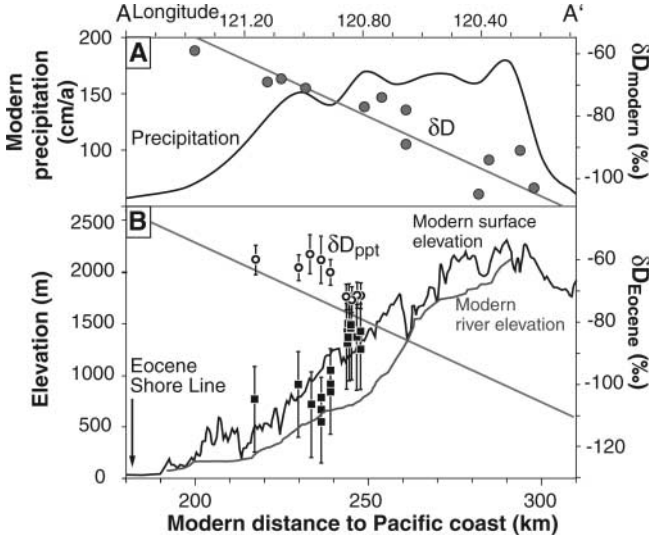


Fig. 3. (**A**) Orographic effect of northern Sierra Nevada as seen in amount and δD of precipitation along modern Yuba river (*23*, *27*). Rainout across modern range correlates with strong decrease in δD_{ppt}. (**B**) A similar correlation between reconstructed Eocene elevation (squares) and calculated δD of Eocene local meteoric water (circles) is observed for the ancient Yuba River. Modern correlation from (**A**) shown for reference.

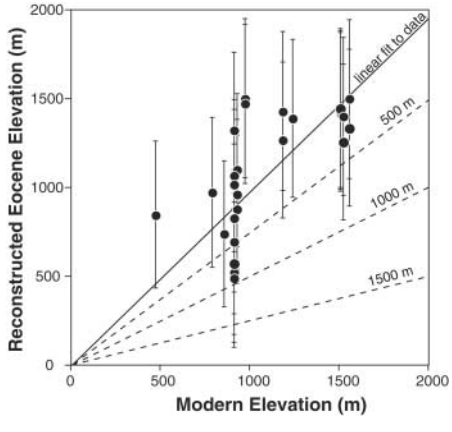


Fig. 4. Modern versus reconstructed Eocene sample elevations for ancestral Yuba River. Modern and Eocene elevations overlap within error, indicating that the Eocene Yuba River attained elevations similar to the interfluves of the modern Yuba river where Eocene river deposits are preserved. Solid line is a fixed–sea level linear regression for the calculated Eocene elevations; dashed lines indicate expected trends for variable amounts of post-Eocene surface uplift.

States following Cretaceous thickening of the lithosphere (*26*). These stable isotope data alone do not preclude Pliocene surface uplift for the northern Sierra Nevada due to mantle delamination as has been proposed for the southern Sierra Nevada (*6*–*8*). In fact, our estimates of post-Eocene uplift are within the range of elevation change predictable for mantle delamination and may actually constrain the magnitude of mantle control on Pliocene surface uplift in the northern Sierra Nevada (Fig. 4). However, they are incompatible with a model of an Eocene low-elevation landscape that experienced substantial (1.5 to 2.0 km) Pliocene surface uplift. Calling upon a Pliocene age for major surface uplift in the northern Sierra Nevada would either require post-Eocene leveling of the mountains with subsequent renewed surface uplift or a tectonic and mechanical decoupling of the northern and southern Sierra Nevada, both of which are unrecognized in the geologic record.

References and Notes

1. P. Molnar, P. England, *Nature* **346**, 29 (1990).
2. C. Beaumont, R. A. Jamieson, M. H. Nguyen, B. Lee, *Nature* **414**, 738 (2001).
3. J. R. Unruh, *Geol. Soc. Am. Bull.* **103**, 1395 (1991).
4. J. Wakabayashi, T. L. Sawyer, *J. Geol.* **109**, 539 (2001).
5. C. H. Jones, G. L. Farmer, J. Unruh, *Geol. Soc. Am. Bull.* **116**, 1408 (2004).
6. G. Zandt *et al.*, *Nature* **431**, 41 (2004).
7. M. N. Ducea, J. Saleeby, *Contrib. Mineral. Petrol.* **133**, 169 (1998).
8. J. Saleeby, M. N. Ducea, D. Clemens-Knott, *Tectonics* **22**, doi:10.1029/2002TC001374 (2003).
9. W. R. Dickinson, R. V. Ingersoll, S. A. Graham, *Geol. Soc. Am. Bull.* **90**, 1458 (1979).
10. B. Wernicke *et al.*, *Science* **271**, 190 (1996).
11. M. A. House, B. P. Wernicke, K. A. Farley, *Nature* **396**, 66 (1998).
12. M. A. Poage, C. P. Chamberlain, *Tectonics* **21**, 1601 (2002).
13. M. K. Clark, G. Maheo, J. Saleeby, K. A. Farley, *GSA Today* **15**, 4 (2005).
14. G. M. Stock, R. S. Anderson, R. C. Finkel, *Earth Surf. Process. Landforms* **30**, 985 (2005).
15. E. E. Small, R. S. Anderson, *Science* **270**, 277 (1995).
16. W. Lindgren, *U.S. Geol. Surv. Prof. Pap.* **73** (1911).
17. W. H. Yeend, *U.S. Geol. Surv. Prof. Pap.* **722** (1974).
18. Available as supporting material on *Science* Online.
19. C. N. Garzione, J. Quade, P. G. DeCelles, N. B. English, *Earth Planet. Sci. Lett.* **183**, 215 (2000).
20. D. B. Rowley, R. T. Pierrehumbert, B. S. Currie, *Earth Planet. Sci. Lett.* **188**, 253 (2001).
21. A. Mulch, C. Teyssier, M. A. Cosca, O. Vanderhaeghe, T. W. Vennemann, *Geology* **32**, 525 (2004).
22. M. A. Poage, C. P. Chamberlain, *Am. J. Sci.* **301**, 1 (2001).
23. N. L. Ingraham, B. E. Taylor, *Water Resources Res.* **27**, 77 (1991).
24. IAEA, http://isohis.iaea.org/userupdate/GNIP2001yearly.xls (2001).
25. C. D. Henry, M. E. Perkins, *Geology* **29**, 719 (2001).
26. P. G. DeCelles, *Am. J. Sci.* **304**, 105 (2004).
27. Spatial Climate Analysis Service, Oregon State University, www.ocs.oregonstate.edu/prism, created April 2005.

28. We acknowledge C. H. Jones and two anonymous reviewers for thorough and constructive reviews. This work was supported by NSF grant EAR-0309011 to C.P.C.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/87/DC1
Materials and Methods
Figs. S1 and S2
Table S1
References and Notes

8 February 2006; accepted 22 May 2006
10.1126/science.1125986

Antibiotic Stress Induces Genetic Transformability in the Human Pathogen *Streptococcus pneumoniae*

Marc Prudhomme,* Laetitia Attaiech,* Guillaume Sanchez, Bernard Martin, Jean-Pierre Claverys†

Natural transformation is a widespread mechanism for genetic exchange in bacteria. Aminoglycoside and fluoroquinolone antibiotics, as well as mitomycin C, a DNA-damaging agent, induced transformation in *Streptococcus pneumoniae*. This induction required an intact competence regulatory cascade. Furthermore, mitomycin C induction of *recA* was strictly dependent on the development of competence. In response to antibiotic stress, *S. pneumoniae*, which lacks an SOS-like system, exhibited genetic transformation. The design of antibiotherapy should take into consideration this potential of a major human pathogen to increase its rate of genetic exchange in response to antibiotics.

Bacterial transformation, originally discovered in the human pathogen *S. pneumoniae* (*1*), relies on a process that is inherent to the species, is independent of extrachromosomal elements, and can be considered the only programmed mechanism for generalized genetic exchange in bacteria. It allows the uptake and integration of exogenous DNA in the recipient genome and is considered to be a form of parasexuality (*2*). Transformation is believed to contribute to the genetic plasticity of *S. pneumoniae* and to play a central role in the adaptation of this pathogen to host defenses (*3*). We sought to establish whether transformation is induced in response to antibiotic stress.

Laboratoire de Microbiologie et Génétique Moléculaires, UMR 5100 CNRS-Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 9, France.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: claverys@ibcg.biotoul.fr

In *S. pneumoniae*, competence for genetic transformation is a transient physiological state allowing efficient DNA uptake (*4*) and a previously unrecognized capacity to kill noncompetent cells (*5*, *6*), a phenomenon referred to as pneumococcal fratricide (*6*) or sobrinicide (*7*). The development of competence requires transcriptional activation of the *com* regulon, which comprises 105 to 124 genes (*8*, *9*), including *recA* (*10*). The RecA protein plays a key role in transformation by catalyzing homologous recombination between the internalized DNA and the recipient genome. The *com* regulon is induced when a competence-stimulating peptide (CSP), encoded by *comC* and exported through a dedicated secretion apparatus ComAB, accumulates in the medium and stimulates its receptor, the membrane-bound histidine kinase ComD (*4*). It is assumed that ComD then autophosphorylates and transphosphorylates its cognate response regulator ComE (*11*), which in turn activates the expression of the so-called early *com* genes (*4*), including *comAB*, *comCDE*, and *comX*. The latter encodes an alternative sigma factor σ^X (*12*), which most probably recognizes a sequence (TACGAATA, hereafter called P*cin*) conserved in the putative promoter regions of the late *com* genes (*4*, *9*). The early control of competence induction is not yet fully understood. It was first suggested that competence induction relies simply on passive CSP accumulation, but we favor an alternative model in which CSP production could be temporarily increased in response to changes in environmental conditions (*4*, *13*). We further propose that competence in *S. pneumoniae* is a general stress response, playing a role similar to that of the SOS response in *Escherichia coli* (*4*).

We tested this hypothesis by investigating the effect of mitomycin C, a DNA-damaging agent known to induce the SOS response, on the *com* regulon. To monitor competence induction (*14*), we used transcriptional fusions with the *luc* gene from the firefly (*Photinus pyralis*) encoding luciferase, the activity of which can be monitored directly in living *S. pneumoniae* cells (*14*). We first used a fusion of *luc* with *ssbB*, a representative of the late *com* genes (*8*, *9*), which encodes a single-stranded DNA–binding protein (*15*). Luciferase activity was monitored during growth with various concentrations of mitomycin C. Expression of the reporter was stimulated by exposure to 25 to 60 ng ml^{-1} of mitomycin C (Fig. 1A and fig. S1A), indicating that *ssbB* was induced when cells were grown in the presence of the DNA-damaging agent. No induction of the same fusion was detected in a *comA* mutant background [that is, in a strain unable to export CSP and therefore to develop spontaneous competence (Fig. 1B)], demonstrating that an

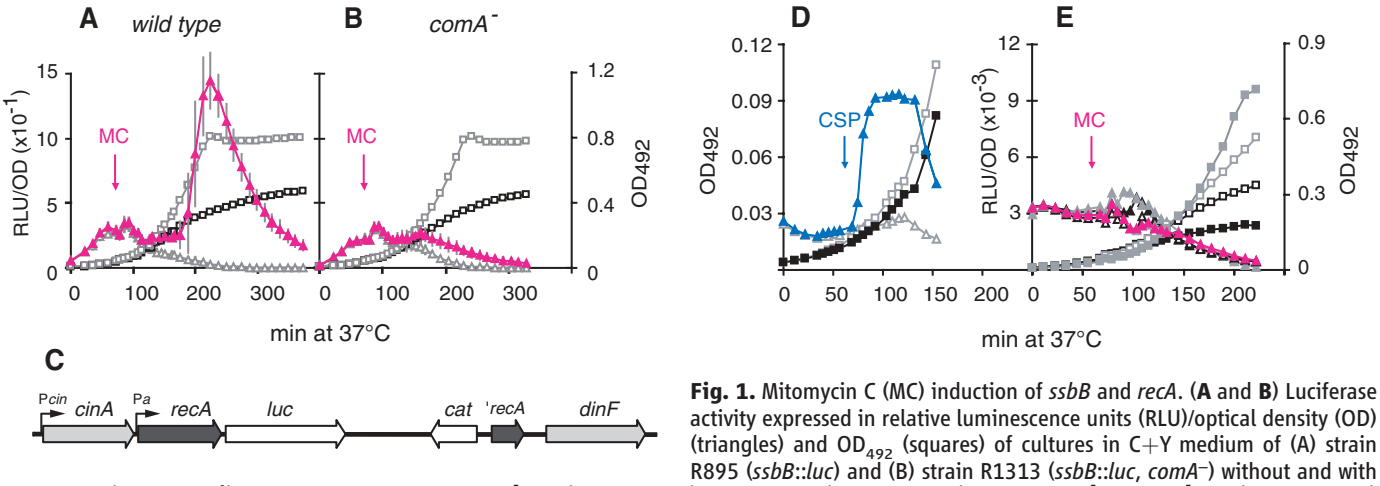


Fig. 1. Mitomycin C (MC) induction of *ssbB* and *recA*. (**A** and **B**) Luciferase activity expressed in relative luminescence units (RLU)/optical density (OD) (triangles) and OD_{492} (squares) of cultures in C+Y medium of (A) strain R895 (*ssbB*::*luc*) and (B) strain R1313 (*ssbB*::*luc*, *comA*⁻) without and with mitomycin C (40 ng ml^{-1}). Curves of luciferase activity [with (red triangles) and without (gray triangles) mitomycin C] and OD [with (black squares) and without (gray squares) mitomycin C] represent compilations of data from 19 and 32 replicate cultures, respectively in (A) and (B). Standard deviations are indicated for luciferase activities only. (**C**) Structure of the *recA* operon with the *recA*::*luc* fusion generated by the integration of plasmid pR432. Plasmid sequences are not drawn to scale. P*cin* and P*a* are indicated by the small branched arrows. (**D** and **E**) Luciferase activity (triangles) and OD_{492} (squares) of cultures in C+Y medium of strain R1624 (*recA*::*luc*, Δ*comC*). In (D), solid blue triangles and solid black squares indicate the presence of CSP (100 ng ml^{-1}), whereas open gray symbols indicate the absence of CSP. In (E), symbols indicate the following concentrations of mitomycin C: solid gray symbols (0 ng ml^{-1}), red triangles and open gray squares (25 ng ml^{-1}), open black symbols (100 ng ml^{-1}), and solid black symbols (400 ng ml^{-1}). Arrows indicate the addition of CSP or mitomycin C after 70 min of incubation.

intact competence regulatory cascade was required for induction by mitomycin C. These data established that a DNA-damaging agent induced the *com* regulon of *S. pneumoniae*.

Mitomycin C is known to trigger the SOS response in *E. coli*. DNA damage caused by mitomycin C blocks the replication fork, generating a single-stranded DNA region to which RecA binds to form a nucleoprotein filament (RecA*) (*16*). The coprotease activity of RecA* then catalyzes the self-cleavage of the SOS repressor LexA (*17*), which leads to induction of the SOS genes, including *recA*. In *S. pneumoniae*, the *recA* gene has been shown to be expressed from two promoters: P*cin*, which generates a 5.7-kb-long transcript in competent cells, and P*a*, a σ^A promoter that directs the synthesis of a 4.3-kb-long transcript (*18*) (Fig. 1C). The competence-specific induction of *recA* (that is, expression from P*cin*) accounts for 95% of the transformation (*10*). To establish whether mitomycin C could induce *recA* expression, a *recA*::*luc* fusion was constructed (Fig. 1C) and validated by measuring its induction with CSP (Fig. 1D). Mitomycin C was found to activate this fusion with kinetics similar to that of *ssbB* (fig. S2). The *recA*::*luc* construct was introduced in a strain unable to develop spontaneous competence because of the deletion of the entire *comC* coding region. The DNA-damaging agent did not induce *luc* expression in this genetic background (Fig. 1E), demonstrating that the induction of *recA* by mitomycin C occurs only from P*cin* and is therefore strictly dependent on the ability of this compound to induce competence. This observation supports the hypothesis that *S. pneumoniae*, which lacks an SOS-like induction mechanism (*7*), instead uses the competence regulatory cascade to coordinate a response to mitomycin C. Induction of the *com* regulon in *S. pneumoniae* and of the SOS response in *E. coli* occurs only after prolonged incubation, at ~2.5 hours (Fig. 1A) and 2 hours, respectively (*19*). The similar delay of unrelated regulatory mechanisms probably reflects the need for a slow accumulation of inducing lesions in both species.

Several antibiotics are known to induce the SOS response in SOS-proficient bacteria (*20*, *21*). To investigate the parallels between competence induction and the SOS response,

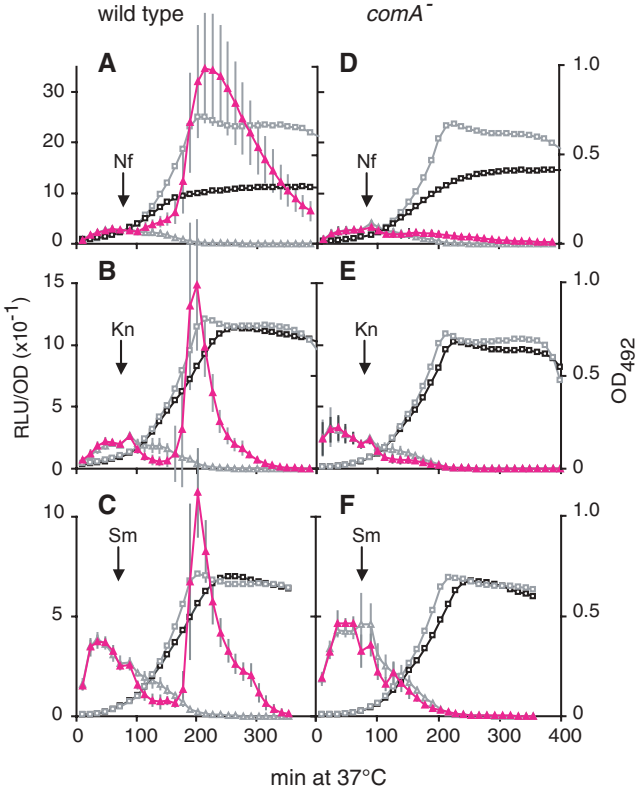


Fig. 2. Antibiotics induce the competence regulon. (**A** to **F**) Luciferase activity (triangles) and OD_{492} (squares) of cultures of strain R895 (*ssbB*::*luc*, wild type) [(A) to (C)], strain R1313 (*ssbB*::*luc*, *comA*⁻) [(D) and (F)], and strain R1047 (E) (*ssbB*::*luc*, *comA*⁻) [(D) to (F)] in C+Y medium with (red and black symbols) and without (gray symbols) antibiotics. Antibiotics were added after 70 min of incubation (arrows) in the following concentrations: 11 μg ml^{-1} of norfloxacin (Nf) [(A) and (D)], 31.25 μg ml^{-1} of kanamycin (Kn) [(B) and (E)], and 12.5 μg ml^{-1} of streptomycin (Sm) [(C) and (F)]. Curves of luciferase activity (with standard deviations) and OD represent compilations of data from 8 cultures [(A) and (D)], 15 cultures [(C) and (F)], or 16 cultures [(B) and (E)] of strains with the respective antibiotics. See Fig. 1 legend for details.

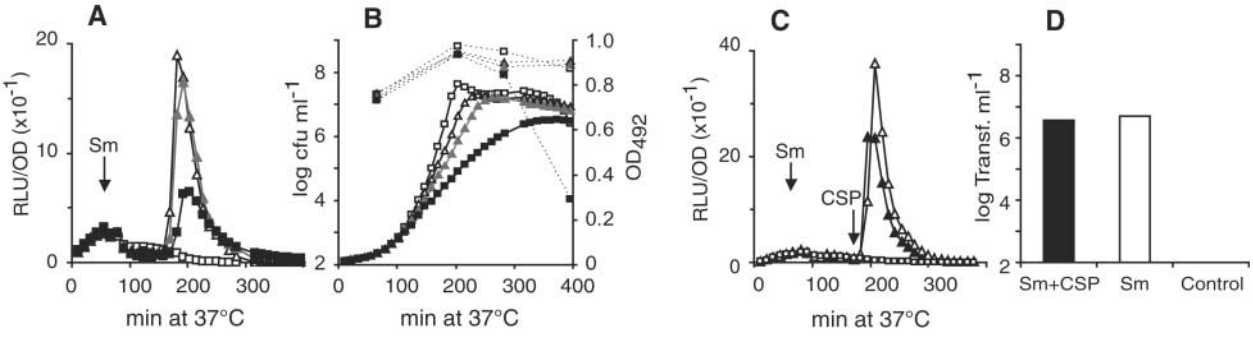


Fig. 3. Induction of genetic transformation by streptomycin. (**A** and **B**) Luciferase activity (A) and growth [(B), OD_{492} and colony-forming units (CFUs) per milliliter] of cultures of strain R895 with 0 (open squares), 6.25 (open triangles), 12.5 (solid gray triangles), or 25 (solid black squares) μg ml^{-1} of streptomycin (Sm). Streptomycin was added after 70 min of incubation [arrow in (A)]. See Fig. 1 legend for details. Cell survival was monitored by plating aliquots [dotted lines in (B)]. (**C**) Luciferase activity of strain R895 grown without (squares) and with (triangles) 6.25 μg ml^{-1} of streptomycin. CSP (100 ng ml^{-1}) was added (right arrow) after 182 min (solid black triangles). (**D**) Genetic transformation in aliquots from cultures in (C) taken after 195 min of incubation and mixed with chromosomal DNA carrying a marker conferring resistance to streptomycin (*14*). The monitoring of transformation with a marker conferring resistance to streptomycin is unrelated to the use of streptomycin to induce competence. The number of streptomycin-resistant chromosomal transformants obtained corresponds to transformation frequencies of 0.65% (Sm) and 0.57% (Sm+CSP).

we measured luciferase synthesis from the *ssbB*::*luc* fusion during growth using a wide range of concentrations of various antibiotics (table S2). Among the protein synthesis inhibitors that were tested, kanamycin and streptomycin triggered competence (Fig. 2, B and C, and fig. S1A), but erythromycin and tetracycline did not (fig. S3, A and B). The fluoroquinolones (norfloxacin, levofloxacin, and moxifloxacin, the latter two of which are used for the treatment of respiratory tract infections), which target type II topoisomerases, DNA gyrase, and topoisomerase IV (*22*), were found to induce the *ssbB*::*luc* fusion (Fig. 2A and fig. S1, B and C). No induction was detected with the DNA gyrase inhibitor novobiocin (fig. S3B), the RNA polymerase inhibitor rifampicin (fig. S3B), the glycopeptide antibiotic vancomycin (fig. S3B), or with the β-lactams ampicillin (fig. S3A) and the third-generation cephalosporin, cefotaxime (fig. S3A). No correlation could be made between the intensity of growth inhibition and the induction of competence (table S2). Similarly to mitomycin C, the induction of *ssbB*::*luc* by aminoglycosides and norfloxacin required an intact competence regulatory cascade, because no induction could be detected in a *comA* mutant (Fig. 2, D to F).

Induction of the *com* regulon normally allows competent cells to take up and integrate exogenous DNA. To test whether antibiotic-induced competence resulted in a bona fide transformation, transforming DNA was added to a culture treated with streptomycin. For the transformation assay, we selected an intermediate concentration of streptomycin (625 ng ml^{-1}), which did not cause severe killing (Fig. 3B). Chromosomal transformants were readily obtained in the streptomycin-treated culture, whereas no transformants were present in the control culture (Fig. 3, C and D). The addition of CSP (100 ng ml^{-1}) to the streptomycin-treated culture did not further increase *ssbB*::*luc* expression (Fig. 3C) or the yield of transformants (Fig. 3D), demonstrating full competence induction by the antibiotic. Chromosomal transformants were also obtained in parallel cultures treated with mitomycin C (60 ng ml^{-1}) and norfloxacin (10 μg ml^{-1}) (fig. S4).

The induction by mitomycin C and fluoroquinolones of the SOS response in *E. coli* and of competence in *S. pneumoniae* suggests that SOS and competence play similar roles in both species. However, the parallel is only partial, because competence was not induced by antibiotics that disrupt cell wall integrity (such as β-lactams), which is contrary to the SOS system in *E. coli* (*21*). Aminoglycosides that induce the *com* regulon (such as kanamycin and streptomycin) did not trigger an SOS response in *E. coli* but instead induced heat-shock protein expression (*23*), suggesting that similar stress signals are processed differently in the two species. Mitomycin C and fluoroquinolones may generate a common signal—chromosome replication arrest—owing to the formation of interstrand cross-links (mitomycin C) or the presence of covalently bound topoisomerases (fluoroquinolones) (*24*). The situation with ribosome inhibitors is more complex, because some such as kanamycin and streptomycin act as competence inducers (Fig. 2, B and C), whereas others such as erythromycin and tetracycline do not (fig. S3, A and B). A parallel can be made with the situation in *E. coli*: Kanamycin and streptomycin, which leave the ribosomal A site empty, induce a heat-shock–like response, whereas erythromycin and tetracycline, which either fill the A site with aminoacyl–transfer RNA (tRNA) or block it, trigger a cold-shock–like response (*23*). The level of ppGpp, which is produced by ribosomes that are stalled by a lack of charged tRNAs, decreases upon the addition of erythromycin or tetracycline (or a reduction in growth temperature) and increases as the growth temperature is increased (*23*, *25*), suggesting that this nucleotide could act as a signal (*4*).

Whatever the underlying mechanism(s) may be, the induction of the *com* regulon by various antibiotics supports the hypothesis that competence is a general stress response of *S. pneumoniae* (*4*, *7*) and is consistent with the proposal that CSP is not an effector of quorum sensing (*7*) but is an alarmone that conveys a stress signal (*4*). As a coordinator of competence, CSP enhances the efficiency of transformation as a rescue process in two ways: by increasing the number of potential transformants (*3*) and by triggering fratricide (*5*) through allolysis (defined as lysis in trans) and the release of DNA from CSP-nonresponsive pneumococcal cells. Whenever these cells differ from the competent population (for example, during cocolonization), allolysis provides a source of genetically diverse DNA. Through the coupling of fratricide and transformation, CSP could thus play a crucial role in generating genetic diversity under stress conditions for a species that seems unable to rely on inducible mutagenic repair (such as the SOS response) (*7*). Consequently, the high incidence of asymptomatic carriage of this pathogen is a major concern, because inappropriate antibiotic treatments could accelerate the occurrence of additional resistant clones and promote the evolution of virulence.

References and Notes

1. F. Griffith, *J. Hyg. (London)* **27**, 113 (1928).
2. J. Maynard Smith, C. G. Dowson, B. G. Spratt, *Nature* **349**, 29 (1991).
3. J. P. Claverys, M. Prudhomme, I. Mortier-Barrière, B. Martin, *Mol. Microbiol.* **35**, 251 (2000).
4. J. P. Claverys, L. S. Håvarstein, *Front. Biosci.* **7**, 1798 (2002).
5. S. Guiral, T. J. Mitchell, B. Martin, J. P. Claverys, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8710 (2005).
6. L. S. Håvarstein, B. Martin, O. Johnsborg, C. Granadel, J. P. Claverys, *Mol. Microbiol.* **59**, 1297 (2006).
7. J. P. Claverys, M. Prudhomme, B. Martin, *Annu. Rev. Microbiol.* (2006).

8. A. Dagkessamanskaia *et al.*, *Mol. Microbiol.* **51**, 1071 (2004).
9. S. Peterson *et al.*, *Mol. Microbiol.* **51**, 1051 (2004).
10. I. Mortier-Barrière, A. de Saizieu, J. P. Claverys, B. Martin, *Mol. Microbiol.* **27**, 159 (1998).
11. E. V. Pestova, L. S. Håvarstein, D. A. Morrison, *Mol. Microbiol.* **21**, 853 (1996).
12. P. Luo, D. A. Morrison, *J. Bacteriol.* **185**, 349 (2003).
13. B. Martin, M. Prudhomme, G. Alloing, C. Granadel, J. P. Claverys, *Mol. Microbiol.* **38**, 867 (2000).
14. To monitor antibiotic-induced competence, cells were incubated in C+Y medium with an initial pH value adjusted so that spontaneous competence induction remained a rare event (see supporting online material).
15. D. E. Grove, S. Willcox, J. D. Griffith, F. R. Bryant, *J. Biol. Chem.* **280**, 11067 (2005).
16. S. C. Kowalczykowski, *Annu. Rev. Biophys. Biophys. Chem.* **20**, 539 (1991).
17. J. W. Little, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1375 (1984).
18. B. Martin, P. García, M. P. Castanié, J. P. Claverys, *Mol. Microbiol.* **15**, 367 (1995).
19. J. W. Beaber, B. Hochhut, M. K. Waldor, *Nature* **427**, 72 (2004).
20. I. Phillips, E. Culebras, F. Moreno, F. Baquero, *J. Antimicrob. Chemother.* **20**, 631 (1987).
21. C. Miller *et al.*, *Science* **305**, 1629 (2004).
22. C. Levine, H. Hiasa, K. J. Marians, *Biochim. Biophys. Acta* **1400**, 29 (1998).
23. R. A. VanBogelen, F. C. Neidhardt, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5589 (1990).
24. K. Drlica, X. Zhao, *Microbiol. Mol. Biol. Rev.* **61**, 377 (1997).
25. M. Cashel, D. R. Gentry, V. J. Hernandez, D. Vinella, in *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, F. C. Neidhardt, Ed. (American Society for Microbiology Press, Washington, DC, 1996), pp. 1458–1496.
26. We thank D. Lane for critical reading of the manuscript. This work was supported in part by European Union grant QLK2-CT-2000-00543.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/89/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 and S2
References

27 March 2006; accepted 1 June 2006
10.1126/science.1127912

Presymptomatic Detection of Prions in Blood

Paula Saá,[1,2] Joaquín Castilla,[1] Claudio Soto[1]*

Prions are thought to be the proteinaceous infectious agents responsible for transmissible spongiform encephalopathies (TSEs). PrP^{Sc}, the main component of the infectious agent, is also the only validated surrogate marker for the disease, and its sensitive detection is critical for minimizing the spread of the disease. We detected PrP^{Sc} biochemically in the blood of hamsters infected with scrapie during most of the presymptomatic phase of the disease. At early stages of the incubation period, PrP^{Sc} detected in blood was likely to be from the peripheral replication of prions, whereas at the symptomatic phase, PrP^{Sc} in blood was more likely to have leaked from the brain. The ability to detect prions biochemically in the blood of infected but not clinically sick animals offers a great promise for the noninvasive early diagnosis of TSEs.

Prion diseases, also called transmissible spongiform encephalopathies (TSEs), are a group of fatal and infectious neurodegenerative diseases, including Creutzfeldt-Jakob disease (CJD) in humans and bovine spongiform encephalopathy (BSE), scrapie, and chronic wasting disease (CWD) in animals. Prions are composed mainly or exclusively of the misfolded prion protein (PrP^{Sc}) (*1*), which replicates in the body, transforming the normal prion protein (PrP^{C}) into more of the misfolded isoform.

Although prion diseases are rare in humans, the established link between a new variant form of CJD (vCJD) and BSE (*2*–*4*) has raised concern about a potential epidemic in the human population. Over the past few years, BSE has become a substantial health problem affecting many countries (*5*), and it seems now apparent that vCJD can be iatrogenically transmitted from human to human by blood transfusion (*6*, *7*). Exacerbating this state of affairs is the lack of a reliable test to identify individuals incubating the disease during the long and silent period from the onset of infection to the appearance of clinical symptoms (*8*–*10*).

PrP^{Sc} is not only the main component of the infectious agent and the most likely cause of the disease, but it is also the only validated surrogate marker for TSEs (*9*). However, PrP^{Sc} concentration is high enough for routine biochemical detection only in the brain and some lymphoid tissues at a time close to the symptomatic stage of the disease (*9*). The development of highly sensitive presymptomatic assays for the biochemical detection of PrP^{Sc} is critical for minimizing the spread of the disease (*9*). One important aim in prion diagnosis is the noninvasive and presymptomatic biochemical detection of PrP^{Sc} in biological fluids, particularly using blood, a fluid known to contain infectivity even before the onset of clinical signs (*6*, *11*, *12*).

PrP^{Sc} has been detected in the blood of sick animals by means of the protein misfolding cyclic amplification (PMCA) technology (*13*). PMCA produces accelerated prion replication, which dramatically amplifies the quantity of PrP^{Sc} present in a sample (*14*, *15*). In a cyclical process, large quantities of PrP^{C} are converted into the misfolded form triggered by the presence of minute and otherwise undetectable amounts of PrP^{Sc}. The method is highly specific for the detection of PrP^{Sc} and leads to a several-million-fold increase in sensitivity as compared to that of standard Western blot assays (*13*).

In order to evaluate the application of PMCA for the detection of prions in blood during the presymptomatic phase, 46 hamsters were inoculated intraperitoneally with 10% brain homogenate of the 263K scrapie strain, and 38 control animals were injected with phosphate-buffered saline (PBS). At different times during the incubation period, groups of animals were killed, blood was collected, and the buffy coat fraction was separated (*13*). Samples of the buffy coat were resuspended directly on healthy hamster brain homogenate and subjected to 144 PMCA cycles. Three different aliquots were tested from each sample. To refresh the substrate, after a round of PMCA cycling, samples were diluted 10-fold into normal brain homogenate, followed by another round of 144 PMCA cycles. This procedure was repeated seven times, because according to our results, this enables the detection of 20 to 50 molecules of monomeric hamster PrP, which seems to correspond to a single unit of infectious oligomeric PrP^{Sc} (*16*).

The first group of hamsters was killed 2 weeks after intraperitoneal inoculation. None of the five infected or control animals showed any detectable quantity of PrP^{Sc} in their blood (Fig. 1 and Table 1). Thus, the PrP^{Sc} present in the inoculum disappeared to undetectable levels during the first few days after inoculation. PrP^{Sc} was, however, readily detectable in blood 1 week later (20 days after inoculation) in 50% of the animals infected but in none of the controls (Fig. 1 and Table 1). The highest percentage of positive animals during the presymptomatic phase was observed 40 days after intraperitoneal inoculation, in which the sensitivity of PrP^{Sc} detection was 60%. After 60 days, the detection of PrP^{Sc} in blood became harder. Indeed, only one out of five animals scored positive at 70 days, whereas none of the five infected hamsters had detectable PrP^{Sc} in their blood 80 days after inoculation (Table 1). At the symptomatic stage, which in this experiment was at 114.2 ± 5.6 days, 80% of animals had PrP^{Sc} in their blood (Fig. 1). We never detected a false positive result in any of the 38 control samples analyzed (Table 1).

The distribution of PrP^{Sc} detection at different times of the incubation period showed

[1]George and Cynthia Mitchell Center for Alzheimer's Disease Research, Departments of Neurology, Neuroscience and Cell Biology, and Biochemistry and Molecular Biology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555–0646, USA. [2]Centro de Biología Molecular, Universidad Autónoma de Madrid, Madrid, Spain.

*To whom correspondence should be addressed. E-mail: clsoto@utmb.edu

an interesting trend (Fig. 2). A first peak of PrP^{Sc} detection was observed early during the presymptomatic phase, between 20 and 60 days after inoculation. The peripheral administration of prions is known to result in an early phase of replication in lymphoid tissues and the spleen, before any infectious material reaches the brain (*17*, *18*). Indeed, little or no infectivity can be detected in the brain of animals peripherally inoculated during the first half of the incubation period (*19*). Thus, it is likely that the source of PrP^{Sc} in blood during the early presymptomatic phase is the spleen and other lymphoid organs. The quantity of PrP^{Sc} in blood goes down after this initial phase and actually disappears 80 days after inoculation (Table 1 and Fig. 2). The rise of PrP^{Sc} in blood during the early presymptomatic phase appears to coincide with the time of its exponential replication in lymphoid organs, whereas the reduction of PrP^{Sc} in blood occurs when infectivity in peripheral tissues has reached a plateau and is migrating from the periphery to the brain (*17*, *18*). Although the explanation for these results in unknown, it is possible that the proportion of circulating lymphocytes carrying PrP^{Sc} is much higher during the exponential phase of peripheral replication than during the stationary phase. At the symptomatic period, PrP^{Sc} can again be detected in the blood of most of the animals (Fig. 2). It has been reported that large quantities of PrP^{Sc} appear in the brain only a few weeks before the onset of clinical signs (*19*, *20*). Thus, PrP^{Sc} in blood samples at the symptomatic stage is likely to have come from brain leakage. It is known that at the time of symptomatic disease, TSE-affected individuals have extensive brain degeneration in the form of massive neuronal death, synaptic alterations, and brain inflammation (*21*). These abnormalities probably cause a disruption of the blood/brain barrier resulting in the leakage of cerebral proteins to the blood (*22*), in particular PrP^{Sc}, which by this time is highly abundant in the brain.

Infectivity studies have shown that the blood carries prions in both the symptomatic and presymptomatic stages of the disease in animals (*11*, *23*, *24*). Upon experimental BSE infection of sheep, infectivity can be transmitted by blood transfusion from asymptomatic infected animals (*25*), indicating that the infectious agent is present in blood during the incubation period. Recently, three cases of vCJD have been associated with blood transfusion from asymptomatic donors who subsequently died from vCJD (*6*, *7*). The alarmingly high proportion of cases transmitted by blood transfusion suggests that prions exist in relatively elevated quantities in the blood of individuals silently incubating vCJD. Based on studies with animal models, it is believed that all of the human population may be susceptible to vCJD infection (*26*), although clinical cases have so far occurred only in methionine homozygotes at codon 129 in the human prion protein gene. Because the incubation period may be several decades, it is currently unknown how many people may be in an asymptomatic phase of

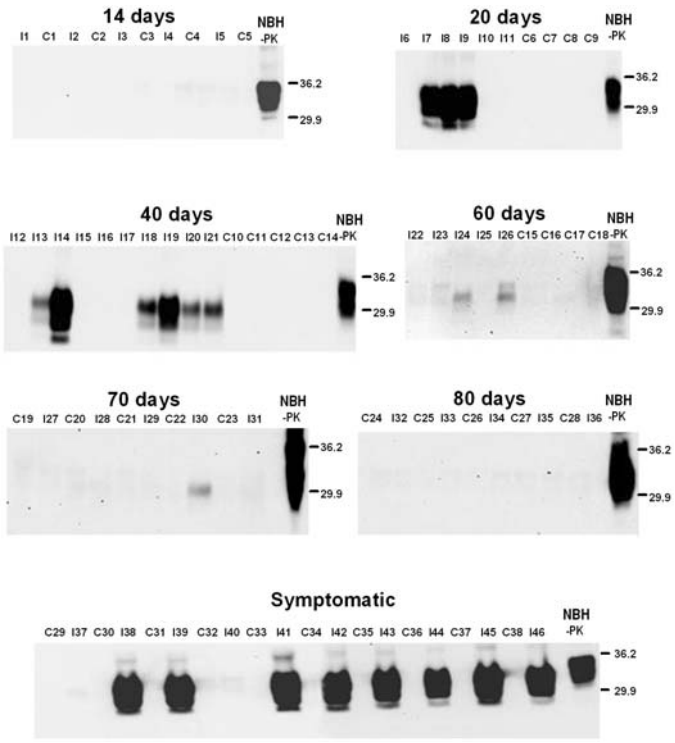


Fig. 1. PrP^{Sc} detection in the blood of scrapie-infected hamsters by PMCA. Blood samples from groups of scrapie-inoculated and control animals were taken at different times during the incubation period. Three milliliters of blood were separated in three aliquots of 1 ml each to prepare the buffy coat (*13*). Samples were subjected to 144 cycles of PMCA. Ten microliters of the sample from this first round of amplification were diluted into 90 μl of normal brain homogenate, and a new round of 144 PMCA cycles was performed. This process was repeated a total of seven times. Each panel represents the results obtained in the seventh round of PMCA with the samples from each group of animals, which are representative of the three independent aliquots taken from each animal. Ix, samples from hamsters infected with 263K scrapie; Cx, samples from control animals injected with PBS. All samples were treated with proteinase K (PK) before electrophoresis, except for the normal brain homogenate (NBH), in which no PK treatment (–PK) is indicated.

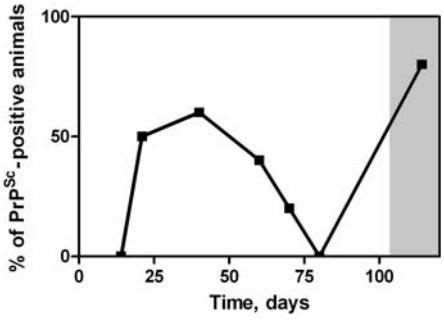


Fig. 2. Proportion of animals whose blood was PrP^{Sc} positive at different times during the incubation period. The percentage of samples scoring positive for PrP^{Sc} in blood is represented versus the time after inoculation at which samples were taken. Two phases of PrP^{Sc} detectability were observed: an early stage during the incubation period, which probably corresponds to the time during which peripheral prion replication in lymphoid tissues is occurring, and a second phase at the symptomatic stage, in which the brain contains extensive quantities of PrP^{Sc}. The vertical gray section indicates the symptomatic phase.

Table 1. Number of animals used and results obtained regarding the presymptomatic detection of PrP^{Sc} in the blood.

Time (days)	Controls (positives/total)	Infected (positives/total)	Sensitivity/ specificity
14	0/5	0/5	0%/100%
20	0/4	3/6	50%/100%
40	0/5	6/10	60%/100%
60	0/4	2/5	40%/100%
70	0/5	1/5	20%/100%
80	0/5	0/5	0%/100%
Symptomatic phase	0/10	8/10	80%/100%

vCJD infection. In addition, it is possible that some infected patients may never develop clinical symptoms but will remain asymptomatic carriers who can potentially transmit the disease to other individuals (*26*, *27*). In the absence of screening tests and effective therapies to treat this disease, a formidable worldwide public health challenge lies ahead to prevent further infections, assess infection rates, and treat infected patients. The ability to detect PrP^{Sc}, the major component of infectious prions, biochemically in the blood of infected but asymptomatic experimental animals will hopefully lead to the development of tests for human blood. Indeed, although technically more challenging, the PMCA technology has been adapted to amplify prions of human origin (*20*). The ability to accurately detect PrP^{Sc} in the presymptomatic stages of vCJD would potentially help to reduce the risk that many more people will be infected by this fatal and terrible disease.

References and Notes

1. S. B. Prusiner, *Proc. Natl. Acad. Sci. U.S.A* **95**, 13363 (1998).
2. S. N. Cousens, E. Vynnycky, M. Zeidler, R. G. Will, R. G. Smith, *Nature* **385**, 197 (1997).
3. J. Collinge, *Lancet* **354**, 317 (1999).
4. M. E. Bruce *et al.*, *Nature* **389**, 498 (1997).
5. R. Bradley, P. P. Liberski, *Folia Neuropathol.* **42** (suppl. A), 55 (2004).
6. C. A. Llewelyn *et al.*, *Lancet* **363**, 417 (2004).
7. A. H. Peden, M. W. Head, D. L. Ritchie, J. E. Bell, J. W. Ironside, *Lancet* **364**, 527 (2004).
8. Q. Schiermeier, *Nature* **409**, 658 (2001).
9. C. Soto, *Nat. Rev. Microbiol.* **2**, 809 (2004).
10. L. Ingrosso, V. Vetrugno, F. Cardone, M. Pocchiari, *Trends Mol. Med.* **8**, 273 (2002).
11. P. Brown, L. Cervenakova, H. Diringer, *J. Lab. Clin. Med.* **137**, 5 (2001).
12. F. Houston, J. D. Foster, A. Chong, N. Hunter, C. J. Bostock, *Lancet* **356**, 999 (2000).
13. J. Castilla, P. Saa, C. Soto, *Nat. Med.* **11**, 982 (2005).
14. G. P. Saborio, B. Permanne, C. Soto, *Nature* **411**, 810 (2001).
15. C. Soto, G. P. Saborio, L. Anderes, *Trends Neurosci.* **25**, 390 (2002).
16. J. R. Silveira *et al.*, *Nature* **437**, 257 (2005).
17. R. H. Kimberlin, C. A. Walker, *J. Comp. Pathol.* **89**, 551 (1979).
18. M. Glatzel, A. Aguzzi, *Microbes Infect.* **2**, 613 (2000).
19. R. H. Kimberlin, C. A. Walker, *J. Gen. Virol.* **67**, 255 (1986).
20. C. Soto *et al.*, *FEBS Lett.* **579**, 638 (2005).
21. J. Castilla, C. Hetz, C. Soto, *Curr. Mol. Med.* **4**, 397 (2004).
22. W. A. Banks, *J. Neurovirol.* **5**, 538 (1999).
23. P. Brown, *Vox Sang.* **89**, 63 (2005).
24. N. Hunter *et al.*, *J. Gen. Virol.* **83**, 2897 (2002).
25. N. Hunter, *Br. Med. Bull.* **66**, 171 (2003).
26. M. T. Bishop *et al.*, *Lancet Neurol.* **5**, 393 (2006).
27. J. W. Ironside, *Haemophilia* **12**, 8 (2006).

28. This research was supported in part by NIH grants AG0224642 and NS049173.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/92/DC1
Materials and Methods
References

21 April 2006; accepted 5 June 2006
10.1126/science.1129051

Prion-Induced Amyloid Heart Disease with High Blood Infectivity in Transgenic Mice

Matthew J. Trifilo,[1] Toshitaka Yajima,[2] Yusu Gu,[2] Nancy Dalton,[2] Kirk L. Peterson,[2] Richard E. Race,[3] Kimberly Meade-White,[3] John L. Portis,[3] Eliezer Masliah,[4] Kirk U. Knowlton,[2]* Bruce Chesebro,[3]* Michael B. A. Oldstone[1]*

We investigated extraneural manifestations in scrapie-infected transgenic mice expressing prion protein lacking the glycophosphatydylinositol membrane anchor. In the brain, blood, and heart, both abnormal protease-resistant prion protein (PrPres) and prion infectivity were readily detected by immunoblot and by inoculation into nontransgenic recipients. The titer of infectious scrapie in blood plasma exceeded 10^7 50% infectious doses per milliliter. The hearts of these transgenic mice contained PrPres-positive amyloid deposits that led to myocardial stiffness and cardiac disease.

In humans and animals, transmissible spongiform encephalopathies (TSEs), or prion diseases, cause neurodegeneration and death following ingestion or experimental inoculation of infected material. Prion diseases are characterized by the conversion of the normal protease-sensitive host prion protein (PrPsen) to a disease-associated protease-resistant form (PrPres). Although prion disease damages the central nervous system (CNS), infectivity and PrPres can be detected within peripheral tissues, including lymphoid organs in humans, sheep, and deer (*1*, *2*), as well as skeletal muscle (*3*), kidney, and pancreas (*4*) of some transgenic rodent models. Despite the toxic effect on the CNS, few if any histopathological changes have been observed at peripheral sites.

Transmission of TSE disease to humans has resulted from cannibalism, contaminated surgical instrumentation, and tainted growth hormone (*5*–*7*). A human disease termed variant Creutzfeldt-Jakob disease (vCJD) has occurred more recently, apparently through the ingestion of bovine spongiform encephalopathy (BSE)–infected cattle products (*8*). Recent evidence suggests that transmission of vCJD between humans may occur through blood transfusion (*9*, *10*), and this conclusion is supported by experimental transmission of BSE between sheep via blood transfusion (*11*). TSE infectivity has been demonstrated in blood by intracerebral-inoculation in mouse, mink, hamster, and goat models (*7*, *12*–*20*). However, infectivity in such cases is low, $\leq 10^2$ 50% infectious doses (ID_{50}) per ml of blood compared to 10^6 to 10^{10} ID_{50}/g in the brain.

Normal prion protein, PrPsen, is expressed primarily as a membrane-bound, glycophosphatydylinositol (GPI)–anchored protein. The role of cellular PrP membrane anchoring in prion disease has been studied in transgenic mice expressing GPI-negative anchorless PrP, which is secreted from cells (*21*). Intracerebral inoculation of these GPI-negative anchorless PrP transgenic (tg) mice with murine scrapie results in scrapie replication and deposition of PrPres within the brain. Although wild-type (WT) mice infected with scrapie usually develop a nonamyloid form of PrPres, in these tg mice the PrPres is primarily in the form of amyloid plaques (*21*). At the same time, these mice do not manifest the clinical and pathologic alterations normally associated with prion disease, thus demonstrating a separation between PrPres amyloid accumulation and clinical CNS disease (*21*). In the brain of these infected tg mice, PrPres was located primarily within and around endothelial cells (*21*) (Fig. 1A), leading to the hypothesis that anchorless PrPres may be secreted in the blood. Here we examined this possibility.

To determine whether PrPres and/or scrapie infectivity was present in blood, four infected tg mice were bled between 450 and 512 days postinfection (dpi) with the RML strain of scrapie. Inoculation of a 1:500 dilution of blood from all four mice induced scrapie in WT (C57BL/6) recipients in ~145 days. In addition, blood of two mice analyzed by serial dilution titration gave titers of $\geq 1.6 \times 10^7$ and $\geq 1.6 \times 10^5$ ID_{50}/ml blood (Table 1).

[1]Viral-Immunobiology Laboratory, Departments of Molecular and Integrative Neurosciences and Infectology, Scripps Research Institute, La Jolla, CA 92037, USA. [2]Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA. [3]Laboratory of Persistent Viral Diseases, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, MT 59840, USA. [4]Departments of Neurosciences and Pathology, University of California, San Diego, CA 92093, USA.

*To whom correspondence should be addressed. E-mail: mbaobo@scripps.edu (M.B.A.O.), bchesebro@niaid.nih.gov (B.C.), kknowlton@ucsd.edu (K.U.K.)

Blood was also tested by immunoblotting for PrPres. At 500 dpi, blood from infected tg mice had detectable PrPres in both plasma and cellular fractions (Fig. 1B). PrPres in the plasma was four times as abundant as in the cellular fraction (Fig. 1, B and F). With 250 to 500 μl of plasma, five out of six additional infected tg mice analyzed between 160 and 500 dpi displayed PrPres (Fig. 1C). PrPres was not detected at 160 dpi but was found in all tg mice at 281 dpi and thereafter.

In these tg mice, similar brain manifestations are induced by all three scrapie strains tested (RML, 22L, and ME7) (*21*). However, between 300 and 500 dpi with strain 22L, PrPres was not detected in blood cells or plasma. Thus, the RML strain appeared to be exceptionally effective in inducing PrPres in the blood of these tg mice.

Next we examined whether PrPres could deposit within extraneural tissues. By immunoblot, PrPres was found in the spleen and heart, but not in the thymus, of infected tg mice at 400 dpi (Fig. 1D). In infected WT mice with clinical disease at the time near death (160 dpi), PrPres was found in the spleen; however, no PrPres was found in the heart or thymus. PrPres was also not detected in the spleen, heart, thymus, or brain of mock-infected tg mice at any times tested (200 to 425 dpi). Thus, accumulation of PrPres in the heart occurred in infected tg mice only.

To determine the timing of PrPres accumulation in the heart, infected tg mice were killed at 100 to 650 dpi, and tissues were examined by immunoblot (Fig. 1E). Infected tg mice had detectable PrPres by 300 dpi, with levels steadily increasing over the 650 days of observation (Fig. 1, E and G). At 600 dpi, the amount of PrPres in the heart tissue was about one-half of that in brain homogenates of RML-infected WT mice (Fig. 1G). Mock-infected tg mice up to 575 dpi (n = 10) and scrapie-infected WT mice up to 160 dpi (n = 10) showed no detectable PrPres in their hearts by immunoblot or by immunohistochemistry. To determine whether scrapie strains differed in PrPres deposition in the heart, we also injected murine scrapie strains 22L and ME7 into groups of tg mice. Both these strains induced significant amounts of PrPres in the heart, which was detected by 350 dpi.

We analyzed infectivity titers of heart tissue by using two tg mice infected with RML scrapie. Whereas mock-infected tg mice and scrapie-infected WT mice had no detectable infectivity in heart, titers in infected tg mice at 480 and 507 dpi exceeded 3.0×10^6 ID_{50} per heart (Table 1). Previously, titers from 2×10^6 to $>4 \times 10^8$ ID_{50} per brain have been observed in infected tg mice (*21*).

We next performed immunohistopathologic studies to characterize the PrPres distribution within cardiac tissue. Immunohistochemical staining revealed the presence of PrPres in heart by 300 dpi. Most PrPres staining was between myocardial cells, often around or in the walls of capillaries (Fig. 2, B to D). Staining of consecutive sections also revealed thioflavin S (Fig. 2, C and D) and Congo red between myocardial cells, indicating amyloid deposition in the corresponding areas. By 500 dpi, PrPres and thioflavin staining was seen throughout the heart (Fig. 2, E and F) and occasionally was observed on or within myocardial cells (Fig. 2G). By light microscopy, myocardial tissue was predominantly normal; however, by Trichrome staining, a few scattered foci of fibrosis were observed, although these did not exceed 1 to 5% of the total heart tissue.

Mice infected with strain 22L also developed similar PrPres amyloid deposition in cardiac tissue, primarily in areas of microvasculature between the cardiac myocytes and around larger blood vessels (Fig. 2H).

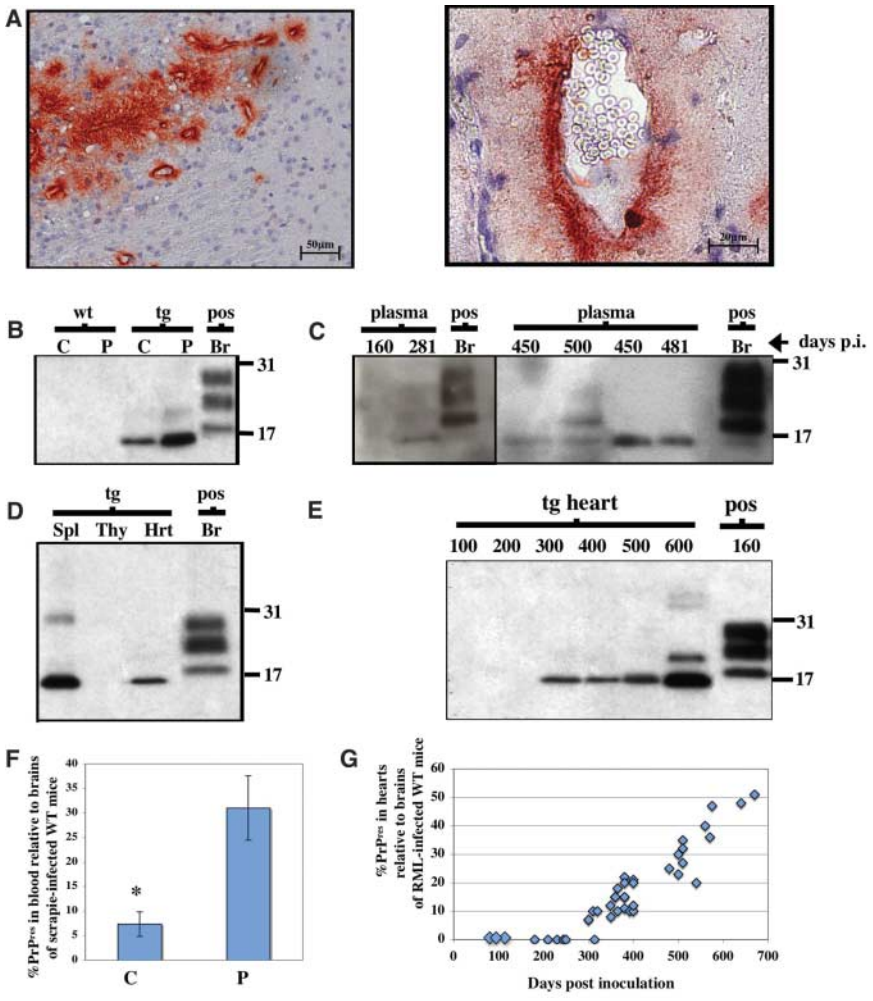


Fig. 1. Detection of PrPres in blood and extraneural tissues of tg mice infected with RML scrapie. (**A**) PrPres lining blood vessels within the corpus callosum of a tg mouse at 500 dpi with scrapie strain RML (PrPres staining in red). (**B** to **E**) PrPres was detected on immunoblots with monoclonal antibody D13. Similar results were observed with R30 antibody. (B) Immunoblot showing PrPres expression in blood cellular (C) and plasma (P) fractions following scrapie infection. For comparison, a sample (pos) equivalent to 10 mg of brain tissue isolated from the brain of a clinically sick RML scrapie-infected C57BL/6 (wt) mouse is shown. (C) Immunoblot showing 250 to 400 μl of plasma from six RML-infected tg mice at times ranging from 160 to 500 dpi. (D) Immunoblot analysis of PrPres in 20 mg of spleen, thymus, and heart from a tg mouse infected for 420 days. Similar results were observed with 10 additional mice. (E) Immunoblot analysis of heart homogenates from individual tg mice at various times after scrapie inoculation. Equivalent results were observed in at least three mice per time point. (**F**) Quantification of PrPres in blood cellular (C) and plasma (P) fractions relative to the amount of PrPres found in the brains of clinically sick WT mice after scrapie infection. Four times as much PrPres was identified in the plasma fraction as in the cellular fraction (*$P \leq 0.05$). (**G**) Quantification of PrPres in the hearts of scrapie-infected tg mice relative to the amount of PrPres in the brains of clinically sick scrapie-infected WT mice. (♦) Each diamond represents data from a single mouse.

Table 1. Titration of scrapie infectivity from heart and blood of RML scrapie-infected GPI-negative anchorless PrP transgenic (tg) and wild-type (WT) mice.

Mice*	Tissue	Dilution†	Incubation period (days)‡	Infected mice/total‡	Titer§
Uninfected tg mouse	Blood	5×10^2	>450	0/6	ND
	Heart	5×10^2	>450	0/6	ND
Infected WT mouse	Blood	5×10^2	>450	0/6	ND
	Heart	5×10^2	>450	0/6	ND
Infected tg mouse #1	Blood	5×10^2	135 ± 10	6/6	$\geq 1.6 \times 10^7$ ID_{50}/ml
	Blood	5×10^3	200 ± 15	3/3	$\geq 1.6 \times 10^7$ ID_{50}/ml
	Blood	5×10^4	250 ± 12	3/3	$\geq 1.6 \times 10^7$ ID_{50}/ml
	Blood	5×10^5	270 ± 18	3/3	$\geq 1.6 \times 10^7$ ID_{50}/ml
	Heart	5×10^2	250 ± 12	3/3	$\geq 3.0 \times 10^6$ ID_{50}/heart
	Heart	5×10^5	347 ± 6	3/3	$\geq 3.0 \times 10^6$ ID_{50}/heart
Infected tg mouse #2	Blood	5×10^2	125 ± 5	6/6	$\geq 1.6 \times 10^5$ ID_{50}/ml
	Blood	5×10^3	205 ± 0	3/3	$\geq 1.6 \times 10^5$ ID_{50}/ml
	Heart	5×10^2	250 ± 14	3/3	$\geq 3.0 \times 10^6$ ID_{50}/heart
	Heart	5×10^5	351 ± 8	3/3	$\geq 3.0 \times 10^6$ ID_{50}/heart

*WT C57BL/6 mice or GPI-negative anchorless PrP tg mice were inoculated intracerebrally with 1×10^6 ID_{50} of RML scrapie. Results illustrate data from a WT mouse killed at the time of severe clinical disease (160 dpi) and from two tg mice killed at 480 dpi (#1) and 507 dpi (#2). One uninoculated control tg mouse was killed at 450 days of age. †Mice were exsanguinated under isoflurane anesthesia, and the heart was removed. Heart homogenates and blood were diluted in sterile phosphate-buffered saline, and 30-μl volumes were inoculated intracerebrally into three to six recipient C57BL/6 mice at the dilutions shown. ‡Recipient C57BL/6 mice were monitored for signs of clinical scrapie including altered gait, ataxia, kyphosis, disorientation, and lethargy. The incubation period was recorded as the time when clinical symptoms were observed and mice were moribund. Scrapie was confirmed by Western blot analysis of PrPres in clinically ill mice. §ID_{50}/heart was calculated by dividing the 50% end-point dilution by the volume inoculated and then multiplying by the weight of the heart (180 mg). Blood titers were calculated as ID_{50}/ml. End-point values were not reached, and the titers shown represent the minimum values possible. ND, not detected.

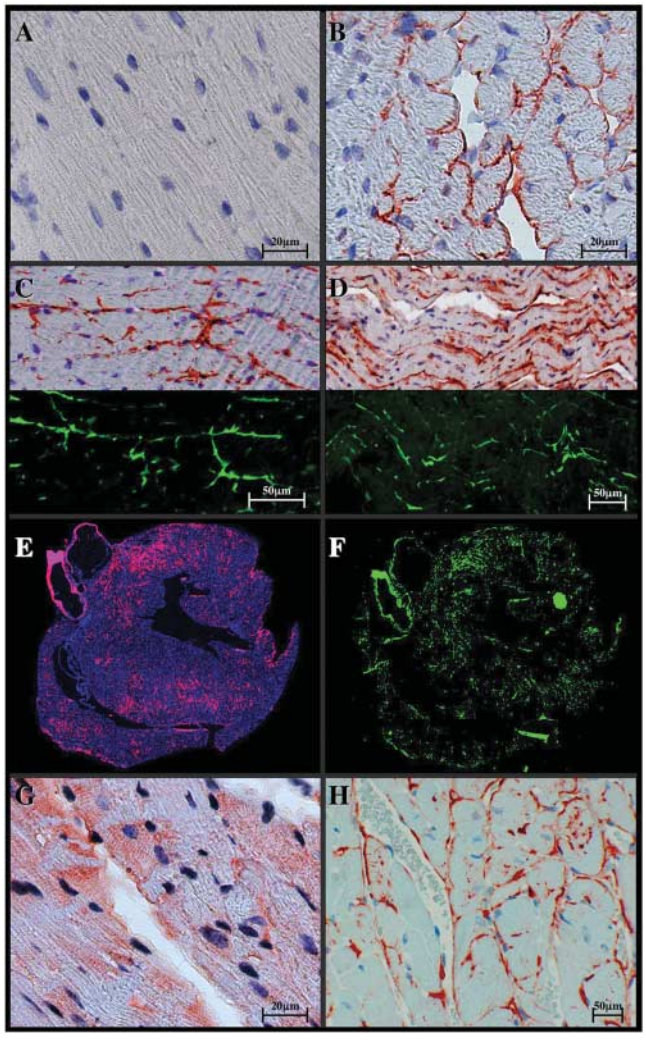


Fig. 2. Detection of PrPres deposits in cardiac samples from scrapie-infected tg mice using PrP monoclonal antibody D13. (**A**) Absence of PrPres in heart tissue in a scrapie-infected clinically ill WT mouse at 160 dpi. Similar results were obtained from an additional 10 scrapie-infected WT mice. (**B**) PrPres (red staining) in heart tissue at 400 days after RML scrapie infection of tg mice. PrPres deposition was observed primarily between myocytes and often around capillaries. (**C** and **D**) Adjacent sections from two mice at 450 dpi with RML scrapie were evaluated for PrPres (red) and amyloid (Thioflavin S; green). PrPres most often localized with amyloid. A similar association between PrPres and amyloid was seen in more than 20 individual hearts from RML scrapie-infected tg mice. (**E**) Detection of widespread deposition of PrPres (pink) and (**F**) amyloid (Thioflavin S; green) from adjacent sections of heart from an infected tg mouse at 500 dpi. (**G**) PrPres could be found in or around myocytes at 500 dpi or later. (**H**) Detection of PrPres between myocardial myocytes and around blood-vessel endothelia in the heart of a mouse infected with 22L scrapie 715 days earlier.

A hallmark of cardiac amyloidosis is a restrictive cardiomyopathy characterized by reduced distensibility or increased stiffness of the heart during diastole. This reduced distensibility is best characterized by the diastolic pressure-volume relation obtained by catheterization of the intact left ventricle. To study possible disordered cardiac function in scrapie-infected tg mice, we compared infected versus uninfected mice by using a series of physiological tests. We found no significant differences in ventricular chamber sizes during end-diastole (left ventricular diastolic diameter) or end-systole (left ventricular systolic diameter), wall thickness (interventricular septum thickness and posterior wall thickness), or fractional shortening, according to echocardiographic analysis (tables S1 and S2). Pressure-volume analysis was performed with catheterization of the left ventricle to determine diastolic and systolic function as described (*22*). Load-independent systolic function was significantly decreased in the infected PrPres tg mice, as demonstrated by a decrease in the slope of the end-systolic pressure-volume relation (Emax) (Fig. 3A). In addition, there was a significant increase in end-diastolic pressure at baseline without an increase in end-diastolic volume in infected tg mice (Fig. 3A). However, the time constant of relaxation, tau, an index for cardiac active relaxation, showed no significant differences between the two groups.

Volume loading (0.5-ml normal saline bolus injection) was used to further assess diastolic distensibility. End-diastolic pressure after volume loading increased in both groups, but remained significantly higher in infected tg mice (Fig. 3B). After volume loading, the preload was altered by transient

occlusion of the inferior vena cava, allowing determination of the end-diastolic pressure-volume relation at the increased volumes. This demonstrated a significant increase ($P \leq 0.05$) in the slope of the end-diastolic pressure-volume relation in the infected tg mice, indicating an increase in stiffness of the left ventricle during diastole (*23*) (Fig. 3B). The increased stiffness and decreased systolic function are characteristic of findings in the early phase of human cardiac amyloidosis (*24*). Thus, the tg mice described in this study appear to have a functional cardiac amyloidosis induced by scrapie infection.

The tg mice used in the present study expressed GPI-negative anchorless PrP, which was secreted from cells (*21*). After scrapie infection, these tg mice produced a soluble form of PrPres found in plasma and blood cells, and the blood contained high levels of scrapie infectivity. The synthesis of secretable anchorless PrP in the infected tg mice is most likely responsible for the accumulation of infectivity and PrPres in the blood. After RML scrapie infection, tg mice produced much greater amounts of blood PrPres than did infected WT mice (*25*), in which 10 million–fold cyclic amplification was required for detection.

Concern about TSE transmission through blood transfusion has grown, and hence the need for reliable tests to screen the human blood supply has become increasingly important. The tg mouse model described in this work reliably produces detectable PrPres and prion infectivity within the blood, so it should be useful for determining the sensitivity of new diagnostic kits and for testing the effectiveness of methods for removal of infectivity.

Evaluation of hearts from infected tg mice indicated the presence of PrPres, amyloid, and scrapie infectivity. Although PrPres was occasionally noted within or on myocardial cells late in infection, most PrPres was found along capillaries in the interstitial spaces. The deposition of amyloid fibrils probably causes the abnormalities in systolic and diastolic ventricular function that limit motion of the ventricular myocytes in this model. However, there was minimal microscopic evidence for myocardial cell damage. These observations are consistent with the modest severity of the cardiac disease in the infected tg mice and suggest that considerable time is required for the phenotype to develop. This is typical of the presentation of amyloid heart disease in humans, in whom diagnosis of the disease before 30 years of age is rare (*23*).

Recently, the deposition of PrPres within a diseased human heart of a 43-year-old CJD patient was reported (*26*). Low amounts of PrPres have also been detected by enzyme-linked immunosorbent assay in heart tissue of monkeys inoculated intracerebrally with vCJD (*27*). Together, these studies suggest that humans and nonhuman primates might occasionally deposit cardiac PrPres, similar to the results obtained with our tg mice. Future human and primate studies should consider more extensive testing for this possibility.

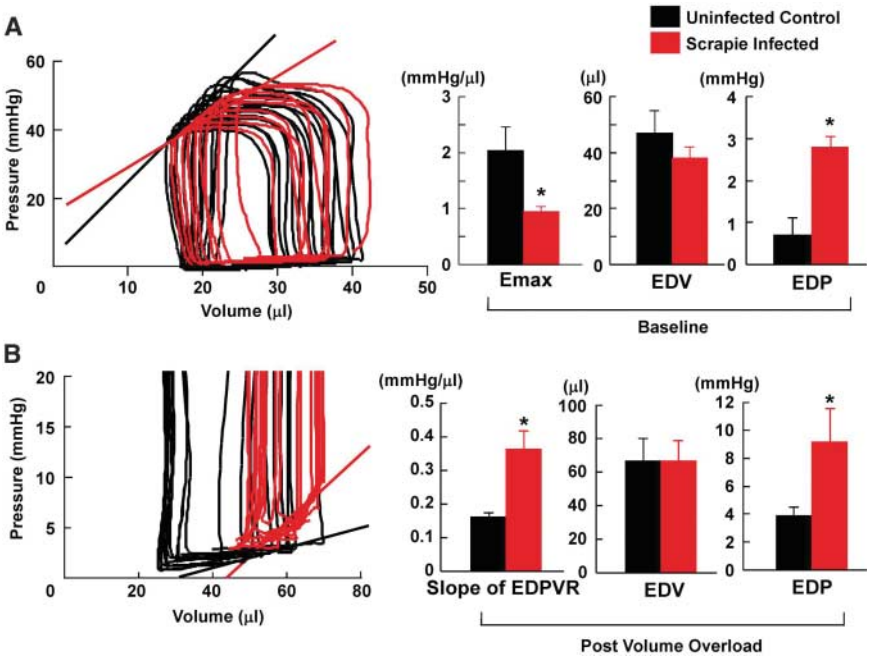


Fig. 3. Hemodynamic assessment of RML scrapie-infected tg mice. Systolic and diastolic function was determined in tg mice with (red, $n = 4$) or without (black, $n = 6$) infection (400 to 500 days old or dpi) by pressure-volume analysis. Measurements are shown for a sequence of beats associated with the decrease in preload induced by transient occlusion of the inferior vena cava [left panel of (A) and (B)]. (**A**) Baseline hemodynamics. Emax, a measure of load-independent systolic function, is defined as the slope of the linear fit line of the end-systolic pressure-volume relation (black and red lines in the left panel). Average baseline indexes for Emax, end-diastolic volume (EDV), and end-diastolic pressure (EDP) are shown. (**B**) Measurement of passive cardiac relaxation (stiffness) after volume load. Cardiac stiffness was further assessed by infusing 0.5 ml of normal saline into the mouse as a volume load. After volume loading, the preload was altered by transient occlusion of the inferior vena cava. The slope of the linear fit of the end-diastolic pressure-volume relation (EDPVR) (black and red lines in the left panel) is proportional to the diastolic stiffness of the left ventricle. Means ± SE for the slope of the EDPVR, EDV, and EDP are shown. *$P \leq 0.05$ as compared to uninfected tg mice.

References and Notes

1. H. Fraser, A. G. Dickinson, *Nature* **226**, 462 (1970).
2. S. B. Prusiner, *Science* **216**, 136 (1982).
3. P. J. Bosque *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3812 (2002).
4. M. Heikenwalder *et al.*, *Science* **307**, 1107 (2005).
5. A. Alperovitch *et al.*, *Lancet* **353**, 1673 (1999).
6. D. C. Gajdusek, C. J. Gibbs Jr., *Res. Publ. Assoc. Res. Nerv. Ment. Dis.* **44**, 254 (1968).
7. C. J. Gibbs Jr. *et al.*, *Science* **161**, 388 (1968).
8. M. R. Scott *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 15137 (1999).
9. C. A. Llewelyn *et al.*, *Lancet* **363**, 417 (2004).
10. A. H. Peden, M. W. Head, D. L. Ritchie, J. E. Bell, J. W. Ironside, *Lancet* **364**, 527 (2004).
11. F. Houston, J. D. Foster, A. Chong, N. Hunter, C. J. Bostock, *Lancet* **356**, 999 (2000).
12. P. Casaccia, A. Ladogana, Y. G. Xi, M. Pocchiari, *Arch. Virol.* **108**, 145 (1989).
13. M. C. Clarke, D. A. Haig, *Vet. Rec.* **80**, 504 (1967).
14. S. C. Collis, R. H. Kimberlin, G. C. Millson, *J. Comp. Pathol.* **89**, 389 (1979).
15. J. Tateishi, Y. Sato, M. Koga, H. Doi, M. Ohta, *Acta Neuropathol. (Berlin)* **51**, 127 (1980).
16. P. Brown *et al.*, *Transfusion* **39**, 1169 (1999).
17. P. Brown *et al.*, *Transfusion* **38**, 810 (1998).
18. L. Cervenakova *et al.*, *Transfusion* **43**, 1687 (2003).
19. K. Holada *et al.*, *J. Virol* **76**, 4649 (2002).
20. Y. Kuroda, C. J. Gibbs Jr., H. L. Amyx, D. C. Gajdusek, *Infect. Immun.* **41**, 154 (1983).
21. B. Chesebro *et al.*, *Science* **308**, 1435 (2005).
22. T. L. Bale *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3697 (2004).
23. W. H. Gaasch, *JAMA* **271**, 1276 (1994).
24. M. R. Zile, D. L. Brutsaert, *Circulation* **105**, 1387 (2002).
25. J. Castilla, P. Saa, C. Soto, *Nat. Med.* **11**, 982 (2005).
26. M. L. Ashwath, S. J. Dearmond, T. Culclasure, *Arch. Intern. Med.* **165**, 338 (2005).
27. C. Herzog *et al.*, *J. Virol.* **79**, 14339 (2005).

28. We acknowledge the technical assistance of A. Andaya, J. Garber, and F. Last. This research was supported by NIH grants AGO4342 (M.B.A.O. and M.J.T.) and 5R01HL66424-04 (K.U.K.) and by NIH training grant NS041219-05 (M.J.T.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/94/DC1
Materials and Methods
Tables S1 and S2
References

11 April 2006; accepted 31 May 2006
10.1126/science.1128635

The Importance of Demographic Niches to Tree Diversity

Richard Condit,[1,2]* Peter Ashton,[3] Sarayudh Bunyavejchewin,[4] H. S. Dattaraja,[5] Stuart Davies,[3] Shameema Esufali,[6] Corneille Ewango,[7] Robin Foster,[8] I. A. U. N. Gunatilleke,[6] C. V. S. Gunatilleke,[6] Pamela Hall,[9] Kyle E. Harms,[10] Terese Hart,[11] Consuelo Hernandez,[12] Stephen Hubbell,[13] Akira Itoh,[14] Somboon Kiratiprayoon,[15] James LaFrankie,[16] Suzanne Loo de Lao,[2] Jean-Remy Makana,[11] Md. Nur Supardi Noor,[17] Abdul Rahman Kassim,[17] Sabrina Russo,[3] Raman Sukumar,[5] Cristián Samper,[18] Hebbalalu S. Suresh,[5] Sylvester Tan,[19] Sean Thomas,[20] Renato Valencia,[12] Martha Vallejo,[21] Gorky Villa,[12] Tommaso Zillio[1,2]

Most ecological hypotheses about species coexistence hinge on species differences, but quantifying trait differences across species in diverse communities is often unfeasible. We examined the variation of demographic traits using a global tropical forest data set covering 4500 species in 10 large-scale tree inventories. With a hierarchical Bayesian approach, we quantified the distribution of mortality and growth rates of all tree species at each site. This allowed us to test the prediction that demographic differences facilitate species richness, as suggested by the theory that a tradeoff between high growth and high survival allows species to coexist. Contrary to the prediction, the most diverse forests had the least demographic variation. Although demographic differences may foster coexistence, they do not explain any of the 16-fold variation in tree species richness observed across the tropics.

Comparative studies of tree demography typically consider the entire community as a unit, ignoring species differences (*1*), simply because most tree inventories include small samples of many species (*2*, *3*). Comparative studies show that tropical forests typically have higher turnover than do temperate forests (*4*) and that higher tree turnover associates with higher tree diversity (*5*). These studies cannot, however, test ecological hypotheses about diversity, coexistence, and demography (*6*–*10*).

A tradeoff between rapid growth and long life span permits species coexistence and can foster diversity: Species reproducing early in life persist despite poor competitive ability by growing rapidly on disturbed sites where resources are abundant. Long-lived species coexist by outliving the weedy invaders, persisting where resources are scarce. This is a familiar and widely known tradeoff in plant and animal communities (*9*–*11*) called the successional-niche hypothesis (*7*, *12*). At a deterministic equilibrium, an indefinite number of species can coexist by this mechanism, each differing from all others along a continuum from short life span (with high growth) to long life span (and low growth). With stochastic demography, however, there is limiting similarity and the equilibrium species richness is finite (*11*, *13*). This hypothesis is widely quoted as an explanation for tropical forest diversity (*14*–*16*). Here, we ask whether species differences along a demographic axis explain why some tropical forests have many more species than others.

If demographic niches are a key force controlling forest diversity, then more diverse forests have more demographic niches. More niches could come about either by spreading demographic rates over a wider range or packing more in the same range. Here, we focus on the first prediction: Tropical forests gain diversity by having a wider range of demographic niches, as reflected by the range of mortality and growth rates across species.

We provide a direct test by quantifying mortality and growth of 4500 tree species in 10 different forests in America, Asia, and Africa (*17*). The 10 sites form a large-scale observation program, spanning a wide range of environmental conditions, designed to provide species-specific information for little-known tropical trees (*18*). At each site, a 20- to 52-ha tree census was set out in extensive, largely undisturbed forest (table S1). Species richness within the census plots differed by 16-fold, from 73 species per 50 ha in a dry forest at Mudumalai, India, to 1167 species per 50 ha in a wet dipterocarp forest in Sarawak, Malaysia (*19*).

Past studies on the demography of individual tree species were based on direct measures of rate constants. These excluded many rare species because their rate estimates are subject to high error (*20*, *21*). To overcome this limitation, we did not simply record species' rates of mortality and growth; instead, we quantified the distribution of demographic rates across the entire community. A hierarchical Bayesian approach accomplishes this with explicit probability models covering both the observations of individual trees within species and the variation among species; all species, including rare ones, are included. For mortality, within-species distributions were modeled with the binomial distribution; for growth, we chose the log-normal based on the tendency for individual growth rates within a species to be highly skewed to the right. By separating within-species variation, the hierarchical model allows focus on the question of how species differ (*10*, *22*, *23*).

At the community level, we had to describe the variation in species' demographic rates across species, and again, we used the log-normal to account for the skew to the right. Histograms of mortality rate m and growth rate g (*24*) are fitted well by the log-normal when rare species are excluded (Fig. 1). The log-normal requires two parameters, μ and σ, the mean and standard deviation of the natural logarithm of m or g, respectively. We were able to estimate values of μ and σ that best describe a community's demography with the use of the Gibbs sampler, simultaneously producing for every species estimates of mortality and growth rates that are adjusted for abundance. That is, for abundant species, the estimate is barely different from the observed rate, but for rare species, it is guided by the community-wide pattern (*25*).

Fitted log-normal distributions for the Lambir forest in Malaysia are plotted with observed histograms of sapling mortality and growth (Fig. 1). The fit is close for more common species (filled bars), demonstrating that the large number of zeroes in the mortality

[1]National Center for Ecological Analysis and Synthesis, 735 State Street, Santa Barbara, CA 93101, USA. [2]Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002–0948, USA. [3]Center for Tropical Forest Science, Asia Program, Arnold Arboretum Asia Program, Harvard University Herbaria, 22 Divinity Avenue, Cambridge, MA 02138, USA. [4]Thai National Park Wildlife and Plant Conservation Department, Research Office, 61 Phaholyothin Road, Chatuchak, Bangkok 10900, Thailand. [5]Center for Ecological Sciences, Indian Institute of Science, Bangalore 560012, India. [6]Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka. [7]University of Missouri, 8001 Natural Bridge Road, St. Louis, MO 63121–4499, USA. [8]Botany Department, The Field Museum, Roosevelt Road at Lake Short Drive, Chicago, IL 60605–2496, USA. [9]Department of Biology, Florida State University, 5051 Quail Valley Road, Tallahassee, FL 32309, USA. [10]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [11]Wildlife Conservation Society, 1725 Avenue Monjiba, Chanic Building, 2nd floor, Ngalinema, Boîte Postale 240, Kinshasa I, Democratic Republic of Congo. [12]Pontifícia Universidad Católica de Ecuador, Post Office Box 17012184, Quito, Ecuador. [13]Department of Plant Sciences, University of Georgia, 2502 Miller Plant Sciences Building, Athens, GA 30602, USA. [14]Plant Ecology Lab, Faculty of Science, Osaka City University, Sugimoto 3-3-138, Sumiyoshi-ku, Osaka 558, Japan. [15]Faculty of Science and Technology, Thammasat University (Rangsit), Klongluang, Patumtani 12121, Thailand. [16]Center for Tropical Forest Science, Asia Program, National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, 637616 Singapore. [17]Forest Environment Division, Forest Research Institute Malaysia, Kepong, Kuala Lumpur 52109, Malaysia. [18]National Museum of Natural History, Smithsonian Institution, 10th Street and Constitution Avenue, NW, Washington, DC 20560, USA. [19]Forest Research Center, Sarawak Forest Department, Km10 Jalun Datak Amar Kalong Ningkan, 93250 Kuching, Sarawak, Malaysia. [20]University of Toronto, 33 Willcock Street, Toronto, Ontario M5S 3B3, Canada. [21]Instituto Alexander von Humboldt, Carrera 13 #28-01, Piso 7, Edifico Palma Real, Bogotá, Colombia.

*To whom correspondence should be addressed: condit@ctfs.si.edu

histogram are sampling artifacts in rare species. Growth rates are also spread by rare species, though not as conspicuously. Fitted distributions for all the forests can be compared graphically (Fig. 2) or with estimated 2.5 and 97.5 percentiles (Table 1). In the supporting online material, tables of mortality and growth rates for all species are provided (database S1).

Most of the forests were dominated by species with sapling mortality rates near 1% year^{-1} (Fig. 2). Even the high-mortality forests, such as Barro Colorado Island (BCI) and Hua Khae Khaeng National Park (HKK), had modes close to 1% year^{-1} and low rates around 0.4% year^{-1}. The main feature separating these high-rate forests from the low-rate sites (such as Pasoh in Malaysia) is the long tail reaching 20% year^{-1} mortality; at Pasoh, nearly all species had mortality rates of <3% year^{-1}. Thus, forests fell broadly in two groups: BCI, HKK, and La Planada had upper sapling mortality rates above 20%, whereas Sinharaja, Lambir, Pasoh, and Yasuni had upper rates below 8%. The Congo sites had exceedingly low mortality stretching to a modest 10% at the upper end.

Distributions of growth rates were similar to distributions of mortality, but growth was about half as variable across species (Fig. 2). Conspicuously, sites with less variation in mortality also had less variation in growth (Fig. 3). These patterns held for larger trees, although mortality and growth rates were lower (table S2 and fig. S1).

Examples from individual species help illustrate. At BCI, a fast-growing understory treelet, *Palicourea guianensis*, had a population of 376 saplings in 1982, and every single one had died by 2005 (nevertheless, the population grew to 851). Although *Palicourea*'s mortality rate is infinite by direct calculation, the Gibbs sampler produces an estimate of 33% year^{-1}. *Alloplectus schultzei*, a small, weedy treelet at La Planada, also suffered 33% year^{-1} mortality, losing 284 of 335 individuals over 6 years. In contrast, of 1161 species at Lambir, none had mortality of 30% year^{-1}, and only two had rates above 20%; at Pasoh between 1987 and 2000, the very highest Gibbs-corrected mortality rate among 802 species was 14% year^{-1}, in *Melastoma malabathricum*.

At the other end of the distribution, *Cupania sylvatica*, a midsized tree of the BCI understory, lost only 10 of 1102 individuals between 1990 and 1995 (0.23% year^{-1}), and *Carapa guianensis* at La Planada, a large and valuable timber tree, lost only 11 of 894 (0.28% year^{-1}). In three census intervals at Pasoh, the lowest

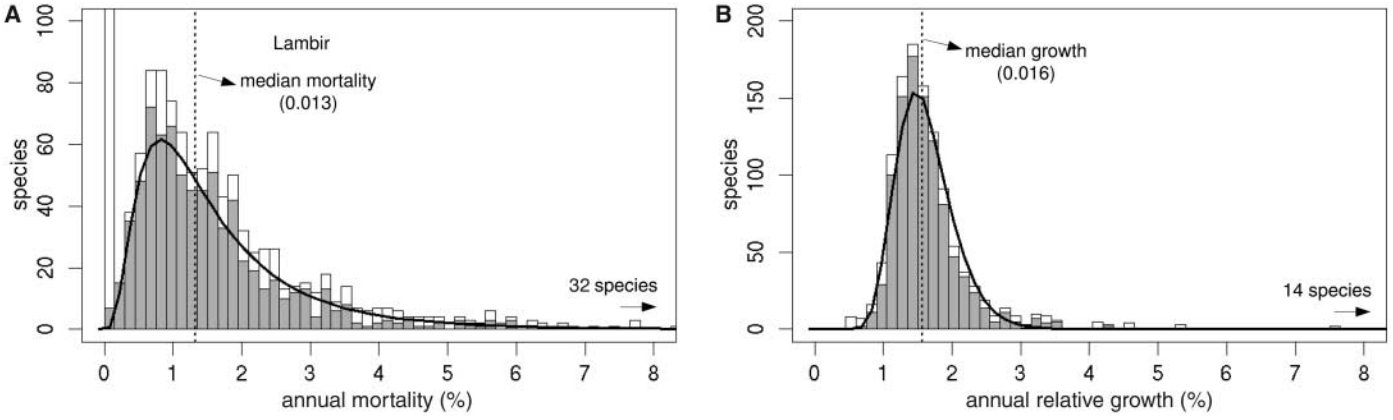


Fig. 1. Distribution of sapling demographic rates of all species in the Lambir plot. (**A**) Annual mortality, *m*, for all individuals with *dbh* = 10 to 99 mm. Filled bars show the histogram of observed mortality rates for the 746 species with ≥50 individuals; open bars add the 415 species with <50 individuals. The open bar at *m* = 0 extends off the graph (162 species had no mortality; 6 of these had <50 individuals). The horizontal axis is *m*, expressed as a percent. The solid line is the fitted log-normal, based on all 1161 species. The dashed vertical line indicates the mean of the logarithm of the fitted distribution (parameter μ, Table 1), which is very close to the median. (**B**) Annual growth, *g*, for individuals 10 to 49 mm in diameter. Filled bars are the histogram for 995 species with ≥10 individuals; open bars for the remaining 154 species. The solid curve and dashed line are the fitted log-normal and the mean of the logarithm, respectively, as in (A). Both histograms are curtailed at 8% to accentuate details where most of the species fall. The number of species above 8% is indicated by arrows.

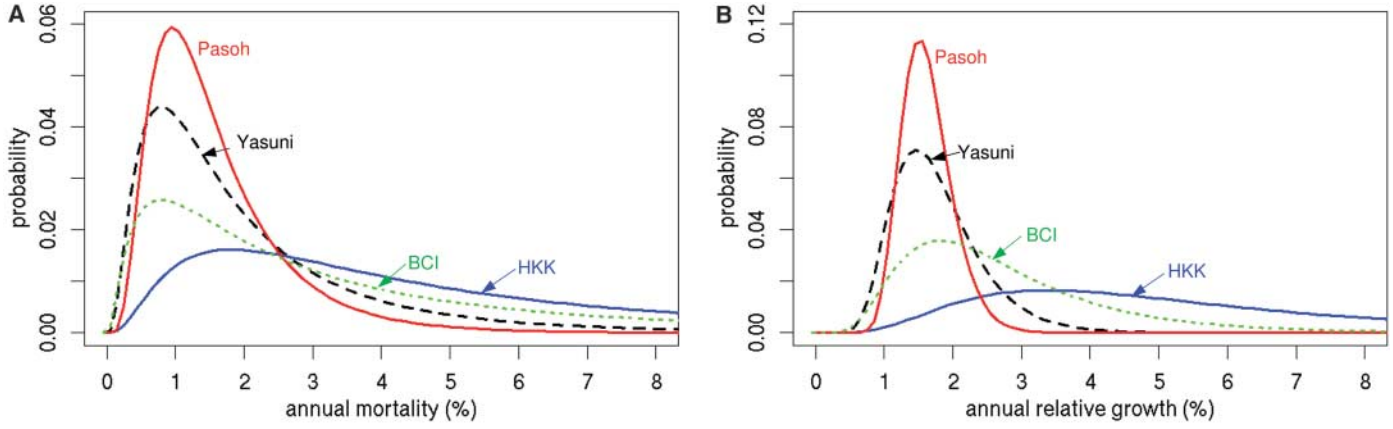


Fig. 2. Comparing the fitted distributions of sapling demography in four forests. (**A**) Annual mortality rate, *m*. (**B**) Annual growth rate, *g*. The lower end of the growth distribution in saplings is limited by measurement accuracy (*30*).

mortality rate recorded was 0.34% year^{-1}, in *Cynometra malaccensis*.

The Mudumalai forest stood out. Saplings had greatly elevated mortality and growth, with rates stretching much higher than any other site. Between 1988 and 1992, every species at Mudumalai had sapling growth of >6% year^{-1}, and only BCI and HKK had many rates this high. At Lenda and Sinharaja, no species grew by 6% year^{-1}. For larger trees (≥100 mm diameter), however, Mudumalai was in line with other forests, having modes of mortality and growth near 1% year^{-1} (table S2). Indeed, trees at Mudumalai had among the lowest rates as well as the highest: *Anogeissus latifolia* had 116 deaths out of 2179 trees from 1988 to 2000, whereas *Kydia calycina* had 1272 of 1328 trees die over the same interval, many because of elephant herbivory (*26*). Their rates differ by 50-fold: 27% year^{-1} in *Kydia* compared with 0.47% year^{-1} in *Anogeissus.*

Three of the sites with long tails of elevated mortality and growth—BCI, HKK, and Mudumalai—have intense annual dry seasons (table S1). Mudumalai and HKK also burn in some years (*26*) (other plots do not burn and none suffer large-scale wind damage). It was not surprising that annual drought elevated mortality. Many species at these sites, however, had exceedingly low rates of mortality and evidently did not suffer much from drought. Conversely, forests lacking the tail of high growth and mortality had no or modest annual dry seasons, including the three forests dominated by Dipterocarpaceae (Sinharaja, Pasoh, and Lambir). Seasonality, however, was not the only factor predicting high variation in demography; the ever-wet cloud forest at La Planada, Colombia, had a wide spread of growth and mortality, comparable to the seasonally dry sites.

Mudumalai and HKK have relatively open canopies compared with all of the other sites, a typical feature of dry forests, and many saplings at Mudumalai are sprouts from large root systems. These are likely reasons for elevated sapling growth at the two sites. In contrast, both BCI and La Planada have dense canopies and dark understories, so canopy openness does not obviously account for the high-growth species at those two sites.

Contrary to the prediction that demographic variability begets species richness, diverse forests had the least variation in demography (Fig. 3 and fig. S1). If anything, the most diverse forests had the fewest demographic niches. At Lambir, high species richness coupled with a low diversity of demographic rates meant that 135 tree species coexisting in close proximity had sapling mortality rates in a narrow window from 0.8 to 1.0% year^{-1}.

We do not question that demographic variability plays some role in species coexistence. In American forests, the familiar genus *Cecropia* is found exclusively in small forest clearings (or outside the forest), where it rapidly

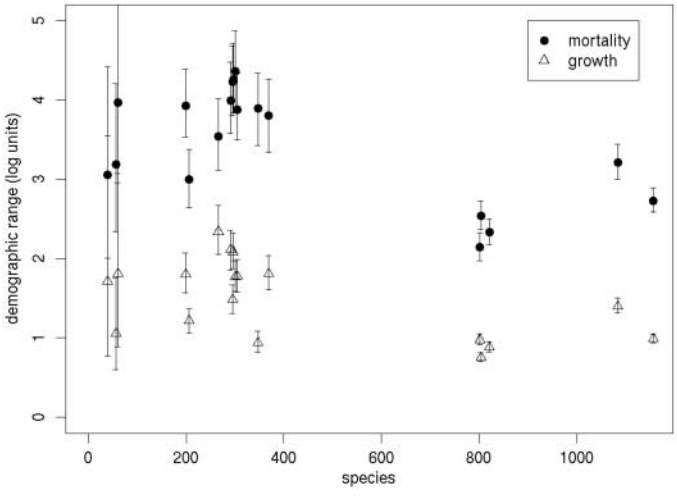


Fig. 3. Range of sapling demographic rates for tree species within a community versus the number of species at the site. The range is the logarithm of the ratio between the 97.5 and 2.5 percentiles of the fitted distributions (Table 1). The range for mortality is given by filled circles; the range for growth is given by open triangles. Sites can be identified by the number of species; for example, Lambir is the most diverse and farthest to the right. Multiple censuses at BCI, Pasoh, and Mudumalai are included, and in each case, fall in a tight group.

Table 1. Variation in sapling mortality and growth rates across species in tropical forests. For mortality, all individuals with *dbh* = 10 to 99 mm were included; for growth, those with *dbh* = 10 to 49 mm were included. Species number refers to those with at least one 10- to 99-mm sapling alive at the outset of a given census interval. Under mortality are percentiles of the distribution of mortality rate parameters (*m*) across species: the median plus lower and upper percentiles (2.5 and 97.5) of the fitted log-normal. Similarly, under growth are percentiles for the distribution of growth rates (*g*) across species. Rates are expressed as percentages (100*m* or 100*g*)—that is, 5 = 5% = 0.05. For each of the percentiles, confidence limits are given, based on the Gibbs sampler (*25*). Information about the sites is presented in table S1.

Site	Years	No. of species	Annual mortality (%)			Relative growth (%)		
			Median	Lower	Upper	Median	Lower	Upper
BCI	82–85	284	3.14 ± 0.46	0.38 ± 0.10	26.0 ± 6.9	2.84 ± 0.16	1.35 ± 0.14	6.0 ± 0.7
BCI	85–90	282	2.56 ± 0.37	0.31 ± 0.08	21.5 ± 5.5	2.41 ± 0.18	0.85 ± 0.12	6.8 ± 1.0
BCI	90–95	282	2.85 ± 0.43	0.32 ± 0.09	25.3 ± 6.7	2.15 ± 0.13	0.89 ± 0.09	5.2 ± 0.7
BCI	95–00	285	3.35 ± 0.42	0.48 ± 0.11	23.3 ± 5.7	1.97 ± 0.12	0.81 ± 0.09	4.8 ± 0.6
BCI	00–05	285	2.91 ± 0.41	0.40 ± 0.10	21.4 ± 5.5	2.10 ± 0.16	0.73 ± 0.10	6.1 ± 0.9
Yasuni	96–03	1077	1.55 ± 0.10	0.31 ± 0.04	7.7 ± 0.9	1.67 ± 0.04	0.83 ± 0.04	3.4 ± 0.2
La Planada	97–03	218	3.22 ± 0.47	0.45 ± 0.13	22.9 ± 5.9	2.30 ± 0.17	0.93 ± 0.13	5.7 ± 0.8
Pasoh	87–90	802	1.04 ± 0.06	0.36 ± 0.04	3.0 ± 0.3	2.25 ± 0.05	1.38 ± 0.05	3.7 ± 0.2
Pasoh	90–95	801	1.35 ± 0.07	0.42 ± 0.05	4.3 ± 0.4	1.59 ± 0.03	1.02 ± 0.04	2.5 ± 0.1
Pasoh	95–00	804	1.69 ± 0.09	0.47 ± 0.06	6.0 ± 0.6	1.55 ± 0.03	1.07 ± 0.03	2.3 ± 0.1
Lambir	92–97	1161	1.32 ± 0.07	0.34 ± 0.03	5.2 ± 0.4	1.57 ± 0.03	0.96 ± 0.03	2.6 ± 0.1
HKK	93–99	256	4.11 ± 0.57	0.70 ± 0.19	24.1 ± 6.6	4.83 ± 0.45	1.53 ± 0.26	15.2 ± 2.7
Mudumalai	88–92	56	13.06 ± 3.48	2.65 ± 1.42	64.4 ± 36	7.87 ± 1.26	4.65 ± 1.52	13.3 ± 3.8
Mudumalai	92–96	52	17.06 ± 6.43	2.35 ± 1.51	124 ± 113	6.35 ± 1.67	2.57 ± 1.38	15.7 ± 11
Mudumalai	96–00	39	7.96 ± 2.70	1.73 ± 1.13	36.6 ± 25	5.71 ± 1.53	2.42 ± 1.69	13.4 ± 9.8
Sinharaja	95–01	205	1.35 ± 0.17	0.30 ± 0.07	6.0 ± 1.3	1.38 ± 0.07	0.75 ± 0.07	2.5 ± 0.2
Edoro	94–00	368	1.43 ± 0.20	0.21 ± 0.06	9.6 ± 2.5	1.41 ± 0.09	0.57 ± 0.07	3.5 ± 0.5
Lenda	94–00	346	1.26 ± 0.19	0.18 ± 0.06	8.8 ± 2.2	1.06 ± 0.04	0.66 ± 0.05	1.7 ± 0.1

colonizes and rapidly dies. The upper end of sapling mortality and growth distributions in America is set by gap specialists: *C. obtusifolia* at BCI (12% year^{-1} mortality, 14% year^{-1} growth), *C. sciadophylla* at Yasuni (5.0% year^{-1} mortality, 6.3% year^{-1} growth), and *C. monostachya* at La Planada (8.8% year^{-1} mortality, 8.2% year^{-1} growth). Diverse Southeast Asian forests lacked species with such high rates (*27*).

The most diverse tropical forests are the least diverse demographically. It remains plausible that demographic niches are packed more tightly in some forests than others, but this seems unlikely, because packing should depend only on population size and turnover, which do not vary much. Moreover, the successional-niche hypothesis is not favored by the strong peak in demographic rates near 1 to 2% year^{-1}; if demographic niches were crucial, then rates ought to be spread evenly over the entire range (*28*). Instead, the similarity in demography of many species suggests trait convergence (*29*). We believe that broad diversity differences are due to the source pool of different biogeographic regions, and that demographic differences play a minor role in species coexistence.

References and Notes

1. M. D. Swaine, D. Lieberman, F. E. Putz, *J. Trop. Ecol.* **3**, 359 (1987).
2. Most forest censuses cover a single hectare or less and include only larger trees. Except in Europe and North America, such small samples capture less than half the local species, with many represented by a single individual.
3. N. Pitman, J. Terborgh, M. R. Silman, P. Núñez V., *Ecology* **80**, 2651 (1999).
4. N. L. Stephenson, P. J. van Mantgem, *Ecol. Lett.* **8**, 524 (2005).
5. O. L. Phillips, P. Hall, A. H. Gentry, S. A. Sawyer, R. Vásquez, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2805 (1994).
6. D. Tilman, *Plant Strategies and the Dynamics and Structure of Plant Communities* (Princeton Univ. Press, Princeton, NJ, 1988).
7. S. W. Pacala, M. Rees, *Am. Nat.* **152**, 729 (1998).
8. B. M. Bolker, S. W. Pacala, C. Neuhauser, *Am. Nat.* **162**, 135 (2003).
9. M. Rees, R. Condit, M. Crawley, S. Pacala, D. Tilman, *Science* **293**, 650 (2001).
10. J. S. Clark, J. Mohan, M. Dietze, I. Ibañez, *Ecology* **84**, 17 (2003).
11. M. B. Bonsall, V. A. A. Jansen, M. P. Hassell, *Science* **306**, 111 (2004).
12. J. S. Clark, S. LaDeau, I. Ibañez, *Ecol. Monogr.* **74**, 415 (2004).
13. D. Tilman, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10854 (2004).
14. P. J. Grubb, *Biol. Rev. Camb. Philos. Soc.* **52**, 107 (1977).
15. S. J. Wright, *Oecologia* **130**, 1 (2002).
16. J. Silvertown, *Trends Ecol. Evol.* **19**, 605 (2004).
17. Table 1 includes 4956 species, but some occur at more than one site, and taxonomy has not been fully aligned across sites. We estimate 4500 different species at the 10 sites.
18. R. Condit, *Trends Ecol. Evol.* **10**, 18 (1995).
19. The Sarawak plot covers 52 ha. For comparison, the number 1167 was taken from a 50-ha subset.
20. J. K. Vanclay, *J. Trop. For. Sci.* **4**, 15 (1991).
21. R. Condit, S. P. Hubbell, R. B. Foster, *Ecol. Monogr.* **65**, 419 (1995).
22. A. Gelman, J. B. Carlin, H. S. Stern, D. B. Rubin, *Bayesian Data Analysis* (Chapman and Hall, Boca Raton, FL, 1995).
23. M. Liermann, R. Hilborn, *Can. J. Fish. Aquat. Sci.* **54**, 1976 (1997).
24. The rate constant m is the derivative of population size N with respect to time, $\frac{dN}{dt}$, due to mortality alone. It is approximated by $\frac{\ln N_2 - \ln N_1}{t_2 - t_1}$. Relative growth rate is $g = \frac{\ln dbh_2 - \ln dbh_1}{t_2 - t_1}$, where *dbh* is the stem diameter at breast height.
25. Materials and methods are available as supporting material on *Science* Online.
26. R. Sukumar, H. S. Suresh, H. S. Dattaraja, N. V. Joshi, in *Forest Biodiversity: Research and Modeling*, F. Dallamier, J. Comiskey, Eds. (Smithsonian, Washington, DC, 1998), pp. 495–506.
27. R. Condit *et al.*, *Phil. Trans. R. Soc. London Ser. B* **354**, 1739 (1999).
28. S. J. Wright, H. C. Muller-Landau, R. Condit, S. P. Hubbell, *Ecology* **84**, 3174 (2003).
29. S. P. Hubbell, *Ecology,* in press.
30. Errors in *dbh* measurements are 1 to 2 mm in saplings, and many saplings show slight negative growth (*25*). We, thus, cannot measure relative growth rates below about 0.5% year^{-1}. Where close to 0.5% year^{-1}, the low end of the sapling growth distributions are probably artifacts. In larger trees, errors are smaller relative to *dbh*, eliminating the problem.

31. Analyses were supported by NSF grant DEB-9806828 of the Research Coordination Network Program to the Center for Tropical Forest Science. Data collection was funded by many organizations, principally the NSF, Andrew W. Mellon Foundation, Peninsula Community Foundation, Smithsonian Tropical Research Institute, Arnold Arboretum (Harvard), Indian Institute of Science, Forest Research Institute of Malaysia, Royal Thai Forest Department, National Institute of Environmental Studies (Japan), and John D. and Catherine T. MacArthur Foundation. We thank the hundreds of field workers who have measured and mapped trees.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1124712/DC1
Materials and Methods
Fig. S1
Tables S1 to S4
References
Database S1
Computer Codes

9 January 2006; accepted 31 May 2006
Published online 8 June 2006;
10.1126/science.1124712
Include this information when citing this paper.

A Single Amino Acid Mutation Contributes to Adaptive Beach Mouse Color Pattern

Hopi E. Hoekstra,[1]* Rachel J. Hirschmann,[1] Richard A. Bundey,[2] Paul A. Insel,[2] Janet P. Crossland[3]

Natural populations of beach mice exhibit a characteristic color pattern, relative to their mainland conspecifics, driven by natural selection for crypsis. We identified a derived, charge-changing amino acid mutation in the melanocortin-1 receptor (*Mc1r*) in beach mice, which decreases receptor function. In genetic crosses, allelic variation at *Mc1r* explains 9.8% to 36.4% of the variation in seven pigmentation traits determining color pattern. The derived *Mc1r* allele is present in Florida's Gulf Coast beach mice but not in Atlantic coast mice with similar light coloration, suggesting that different molecular mechanisms are responsible for convergent phenotypic evolution. Here, we link a single mutation in the coding region of a pigmentation gene to adaptive quantitative variation in the wild.

The identification of the specific molecular changes underlying adaptive variation in quantitative traits in wild populations is of prime interest (*1*, *2*). Pigmentation phenotypes are particularly amenable to genetic dissection because of their high heritability and our knowledge of the underlying developmental pathway (*3*). In a series of classic natural history studies (*4*, *5*), Sumner documented pigment variation in *Peromyscus polionotus*, including eight extremely light-colored "beach mouse" subspecies, which inhabit the primary dunes and barrier islands of Florida's Gulf and Atlantic coasts (*6*). This light color pattern is driven by selection for camouflage (*7*, *8*) as major predators of *P. polionotus* include visual hunters (*9*). Because the barrier islands on the Gulf Coast are <6000 years old (*10*), this adaptive color variation may have evolved rapidly.

We examined the contribution of the melanocortin-1 receptor gene (*Mc1r*) to this adaptive color patterning. MC1R, a G protein–coupled receptor, plays a key role in melanogenesis by switching between the production of dark eumelanin and light pheomelanin (*11*). Mutations in *Mc1r* have been statistically associated with Mendelian color polymorphisms in several mammalian species (e.g., *12–14*) and in natural variants of avian plumage (*15*, *16*).

[1]Division of Biological Sciences and [2]Department of Pharmacology, University of California, San Diego, La Jolla, CA 92093, USA. [3]*Peromyscus* Genetic Stock Center, Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA.

*To whom correspondence should be addressed. E-mail: hoekstra@ucsd.edu

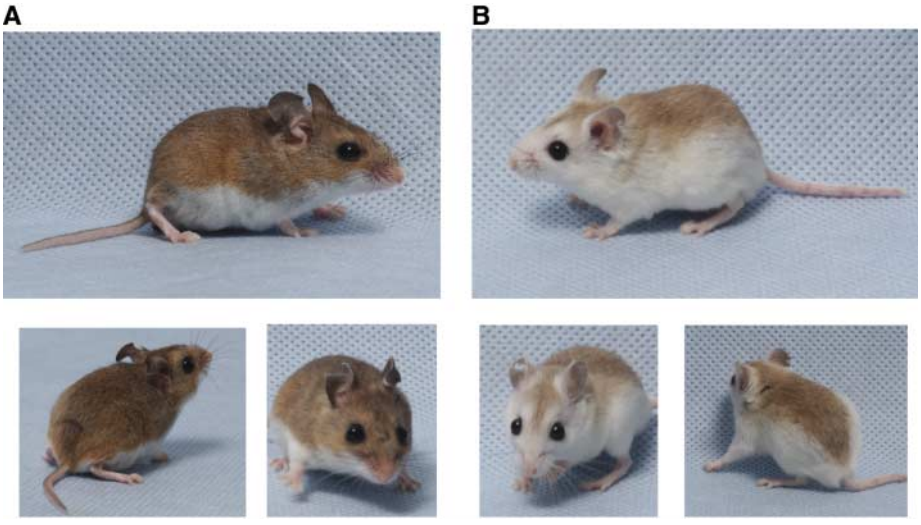


Fig. 1. Photographs of a typical (**A**) mainland mouse (*Peromyscus polionotus subgriseus*) and (**B**) Santa Rosa Island beach mouse (*P. p. leucocephalus*), highlighting the color pattern differences in the flank, face, and rump between these subspecies.

Table 1. Statistical association between allelic variation at *Mc1r* and seven pigmentation traits in 459 F_2 individuals. Phenotypes were categorically scored based on the pigmentation pattern on individual hairs in seven areas (traits). For all seven traits, each F_2 was scored as follows: 0, unpigmented hair; 1, partially pigmented hair; and 2, fully pigmented hair. The total number of F_2s with each phenotypic score is partitioned by *Mc1r* genotype for each trait: RR, homozygous for the dark *Mc1r* allele; RC, heterozygous; and CC, homozygous for the light *Mc1r* allele. The frequency of F_2s with each phenotype score is shown for each trait, with the phenotype of all F_1 individuals in bold. The mean phenotypic score (±SD) for each trait is provided and is also partitioned among each genotype. The percentage of variance explained by *Mc1r* genotype (PVE) and likelihood ratio chi-square (X^2) values are shown. *All seven tests were statistically significant ($P < 0.0001$).

Trait	*Mc1r* genotype	Phenotypic score			Mean ± SD	PVE (%)	X^2
		0	1	2			
Whisker		0.37	**0.30**	0.33	0.97 ± 0.84	11.5	115.3*
	RR	12	49	65	1.42		
	RC	69	77	69	1.00		
	CC	85	13	20	0.45		
Rostrum		0.11	0.18	**0.71**	1.61 ± 0.68	36.4	131.4*
	RR	0	0	126	2.00		
	RC	0	40	175	1.81		
	CC	49	42	27	0.81		
Cheek		0.14	**0.60**	0.26	1.12 ± 0.63	27.2	232.9*
	RR	0	91	35	1.28		
	RC	0	159	55	1.25		
	CC	66	23	29	0.69		
Eyebrow		0.16	0.25	**0.59**	1.42 ± 0.76	16.2	141.4*
	RR	0	21	105	1.83		
	RC	18	62	135	1.54		
	CC	56	31	31	0.79		
Ear		0.44	**0.36**	0.20	0.75 ± 0.77	10.2	98.0*
	RR	28	53	45	1.13		
	RC	89	79	47	0.80		
	CC	85	33	0	0.28		
Ventrum		**0.69**	0.11	0.20	0.51 ± 0.81	16.3	122.6*
	RR	55	27	44	0.91		
	RC	144	24	47	0.55		
	CC	118	0	0	0.00		
Ankle		0.45	**0.35**	0.20	0.75 ± 0.77	9.8	93.4*
	RR	24	70	32	1.06		
	RC	94	66	55	0.82		
	CC	88	26	4	0.29		

We sequenced the entire coding region of *Mc1r* [954 base pairs (bp)] in five Santa Rosa Island beach mice (*P. p. leucocephalus*) and five mainland mice (*P. p. subgriseus*) from colonies derived from wild populations (*17*). A single, fixed nucleotide polymorphism (SNP) occurs between the two subspecies, resulting in a charge-changing amino acid variant ($R^{65}C$) in the first intracellular loop of MC1R (fig. S1). To determine the ancestry of this mutation, we genotyped this SNP in 14 other *Peromyscus* species ($N = 45$ individuals) (table S1). These additional species are all fixed for R^{65} as in the mainland "dark allele," suggesting that the "light allele" is derived and not present in any other fully pigmented *Peromyscus* species.

To examine independently the relationship between this mutation ($R^{65}C$) and the derived light-colored phenotype, we generated a large reciprocal F_2 intercross between the Santa Rosa Island beach mouse and the mainland subspecies (Fig. 1). We characterized color phenotype for seven pigmentation traits, which provide an overall description of the continuous variation in color pattern, and genotyped the *Mc1r* allele in 459 F_2 individuals [126 with two dark *Mc1r* alleles (RR), 215 with one dark and one light allele (RC), and 118 with two light alleles (CC)]. Based on the observed phenotypic variation among F_2s, beach mouse color pattern has a multigenic architecture (Table 1). Pairwise correlation (R^2) between traits ranged from 0.147 to 0.500 ($P < 0.05$ for all comparisons), revealing a shared genetic basis among traits but also indicating the role of loci expressed in distinct spatial regions.

Two lines of evidence suggest that both dominant and recessive alleles contribute to the adaptive light color phenotype. First, all F_1 hybrids are intermediate in overall color pattern, with some traits resembling the mainland parent and other traits the beach mouse parent (Table 1). Second, dominance varies at the *Mc1r* locus itself. Phenotypic scores for pigmentation traits among F_2s are sometimes above the mean score of 1, suggesting that the light *Mc1r* allele is partially recessive, and sometimes below the mean score, suggesting that the light *Mc1r* allele is partially dominant (Table 1).

We found a significant statistical association between the $R^{65}C$ polymorphism and all seven traits, although the percentage of phenotypic variance explained by *Mc1r* varied (9.8% to 36.4%) (Table 1). For some pigmentation traits, the association between phenotype and *Mc1r* genotype was notable: All RR F_2s ($N = 126$) expressed the darkest phenotype on the rostrum and, conversely, only CC individuals expressed the lightest rostrum phenotype ($N = 49$). We calculated principal component scores for the combined color traits and evaluated whether there were

significant differences between the *Mc1r* alleles for PC1 scores (PC1 explained 71% of the variance in color traits); these analyses suggest that *Mc1r* accounts for 26% of PC1 ($P < 0.0001$). Together, these data indicate that *Mc1r* is a major effect locus on color pattern, having pleiotropic effects on pigmentation in spatially diverse areas of the body.

Functional tests of the light and dark *Mc1r* alleles demonstrate that the $R^{65}C$ amino acid mutation contributes to variation in adaptive pigmentation. We performed in vitro assays on HEK293T cells expressing light or dark *Peromyscus* coding region alleles at similar levels. Stimulation with an MC1R agonist (NDP-αMSH) and subsequent measurement of generated cyclic adenosine monophosphate (cAMP) by radioimmunoassay revealed that compared with dark receptor–expressing cells, cells that expressed the light MC1R had a statistically significant reduction in basal and stimulated cAMP formation (Fig. 2A) ($P < 0.05$, Student's t test, $n = 4$; maximum responses: dark allele = 120 ± 15 pmol/mg protein, light allele = 28 ± 5 pmol/mg protein). MC1R-mediated cAMP formation is associated with eumelanin production (*18*); thus, reduced cAMP production may underlie the lack of dark pigmentation observed in mice expressing the light MC1R.

The light MC1R displays lower affinity binding to αMSH, and the decreased response to guanine nucleotide by the light receptor implies an altered ability to interact with G proteins (Fig. 2B). Specifically, whereas the dark receptor displays a substantial GppNHp-promoted shift in IC_{50} (from 1.37 ± 0.01 nM in the absence to 13.5 ± 0.1 nM in the presence of GppNHp), the light allele has a decreased affinity for nucleotide and a nonsignificant IC_{50} shift (without GppNHp, 222 ± 30 nM; with, 178 ± 14 nM). These results demonstrate that a single amino acid mutation in MC1R is responsible for reduced ligand binding and G protein coupling, consistent with the reduced ability of the light MC1Rs to promote cAMP formation in vivo.

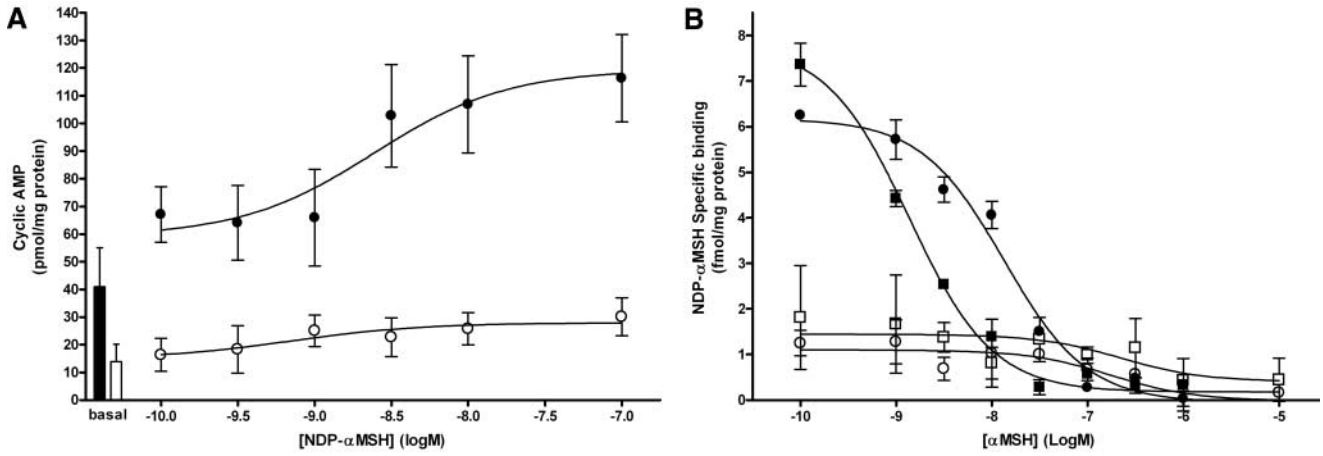


Fig. 2. Cyclic AMP and ligand-binding assays showing functional differences between light and dark *Mc1r* alleles. All data points are means ± SEM. (**A**) HEK293T cells transfected with dark MC1R (filled circles) or light MC1R (open circles) have different cAMP responses to NDP-αMSH stimulation. Bars indicate basal level of cAMP production in cells transfected with dark MC1R (black) and light MC1R (white); lines indicate standard error. (**B**) αMSH competition for [^{125}I]NDP-αMSH binding to HEK293T cells transfected with dark MC1R (filled symbols) or light MC1R (open symbols). Squares indicate assays without GppNHp and circles with GppNHp (100 μM).

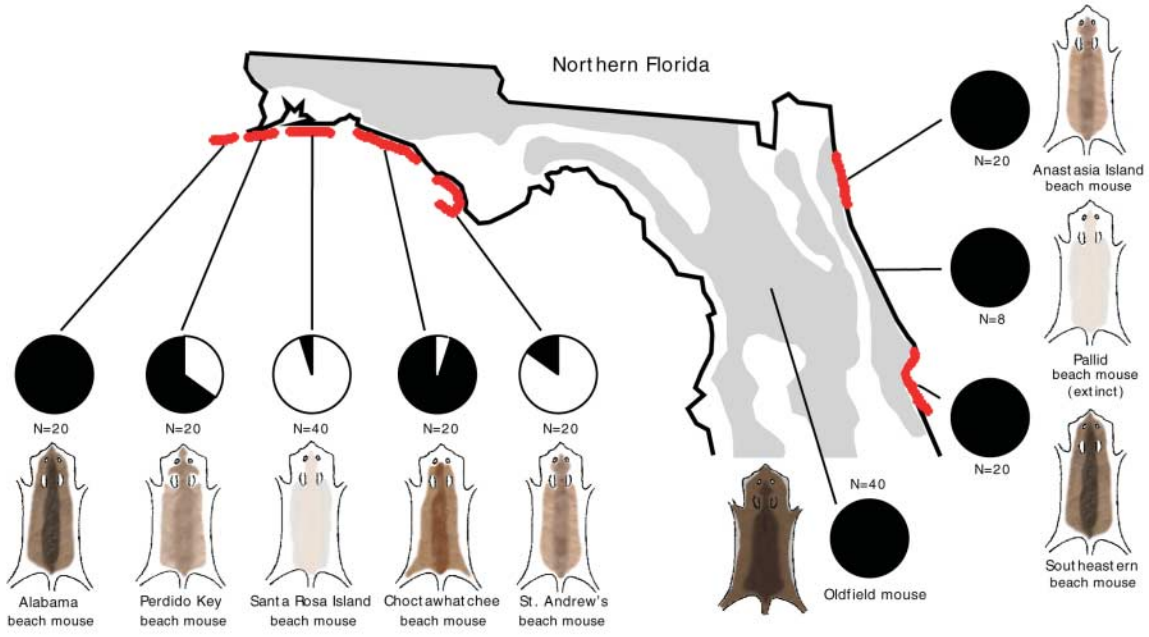


Fig. 3. Frequency of *Mc1r* alleles in one mainland and eight beach mouse subspecies from northern Florida. Images represent typical color patterns for each subspecies. Red areas represent the distribution of each beach mouse subspecies, and the gray area represents the range of mainland subspecies in Florida. Lines indicate sampling locales. Circles represent frequencies of the light *Mc1r* (white) and dark *Mc1r* (black) alleles identified. The number of alleles sampled is provided.

We next examined the frequency of *Mc1r* alleles in natural populations by genotyping the informative SNP in eight beach mouse subspecies and one mainland subspecies (Fig. 3 and table S2). In the Gulf Coast the light allele was at highest frequency (0.95) in the palest beach mouse subspecies, absent in the darkest subspecies, and at intermediate frequencies in the other subspecies (0.05 to 0.85) (Fig. 3). The distribution of *Mc1r* allele frequencies among Gulf Coast beach mice is not correlated with geographic distance. The light *Mc1r* allele was not detected in the mainland subspecies (0 of 40 alleles). In natural populations, different combinations of functionally distinct *Mc1r* alleles may contribute to variation in color patterning between mainland and Gulf Coast beach mouse subspecies, among the Gulf Coast subspecies, and within subspecies.

The derived *Mc1r* allele was absent from the similarly light-colored Atlantic coast beach mouse populations. We surveyed *Mc1r* allelic variation in two extant subspecies of beach mice and the extinct pallid beach mouse subspecies (~300 miles east of the nearest Gulf Coast beach mouse population) (Fig. 3). The absence of the derived *Mc1r* SNP from all three Atlantic coast subspecies suggests that the same mutation does not contribute to convergent phenotypic adaptation and that light coloration has evolved independently on the Atlantic coast.

This work has specific implications for understanding the evolutionary mechanisms responsible for adaptive phenotypic change. First, the identification and functional characterization of a single amino acid mutation's effect on quantitative variation provides a convincing exception to a growing number of examples demonstrating that variation in morphology is governed by changes in gene regulatory regions (*19*, *20*). Second, the observation that different combinations of alleles can produce similar pigmentation patterns suggests that distinct molecular mechanisms can underlie adaptive convergence even in similar selective environments (but see *21*). Finally, *Mc1r* represents a large effect locus, containing a quantitative trait nucleotide (QTN) contributing to variation in fitness, consistent with the view that adaptation may often proceed by large steps.

References and Notes

1. H. A. Orr, J. A. Coyne, *Am. Nat.* **140**, 725 (1992).
2. T. F. C. Mackay, *Annu. Rev. Gen.* **35**, 303 (2001).
3. I. J. Jackson, *Annu. Rev. Gen.* **28**, 189 (1994).
4. F. B. Sumner, *Proc. Natl. Acad. Sci. U.S.A.* **15**, 110 (1929).
5. F. B. Sumner, *Proc. Natl. Acad. Sci. U.S.A.* **15**, 481 (1929).
6. W. W. Bowen, W. D. Dawson, *J. Mamm.* **58**, 521 (1977).
7. M. C. Belk, M. H. Smith, *J. Mamm.* **77**, 882 (1996).
8. D. W. Kaufman, *J. Mamm.* **55**, 271 (1974).
9. J. L. VanZant, M. C. Wooten, *Biol. Conserv.* **112**, 405 (2003).
10. F. S. McNeil, *U.S. Geol. Surv.* **221**, 59 (1950).
11. G. S. Barsh, *Trends Gen.* **12**, 299 (1996).
12. E. Eizirik *et al.*, *Curr. Biol.* **13**, 448 (2003).
13. M. W. Nachman, H. E. Hoekstra, S. L. D'Agostino, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5268 (2003).
14. K. Ritland, C. Newton, H. D. Marshall, *Curr. Biol.* **11**, 1468 (2001).
15. N. I. Mundy *et al.*, *Science* **303**, 1870 (2004).
16. E. Theron, K. Hawkins, E. Bermingham, R. E. Ricklefs, N. I. Mundy, *Curr. Biol.* **11**, 550 (2001).
17. Materials and methods are available as supporting material on *Science* Online.
18. L. S. Robbins *et al.*, *Cell* **72**, 827 (1993).
19. S. B. Carroll, *PLoS Biol.* **3**, 1159 (2005).
20. S. B. Carroll, *Endless Forms Most Beautiful: The New Science of Evo-Devo* (Norton, New York, 2005).
21. P. F. Colosimo *et al.*, *Science* **307**, 1928 (2005).

22. We thank M. Dewey for help maintaining mice, and L. Mullen, L. Turner, and C. Yamada for contributions to the molecular work. Natural population samples were provided by M. Ashley, C. Parkinson, A. Suazo, and M. Wooten. Comments by the Hoekstra Lab, M. Hofreiter, J. Kohn, K. Peichel, T. Price, H. Roempler, J. Stinchcombe, and anonymous reviewers improved this manuscript. J.P.C. is supported by NIH-P40-RR14279 and NSF-DBI-0444165. This research is funded by NSF-DEB-0344710 to H.E.H. GenBank Accession Numbers: DQ482848, DQ482850.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/101/DC1
Materials and Methods
Fig. S1
Tables S1 and S2
References

13 February 2006; accepted 27 April 2006
10.1126/science.1126121

Decay of Endoplasmic Reticulum-Localized mRNAs During the Unfolded Protein Response

Julie Hollien and Jonathan S. Weissman*

The unfolded protein response (UPR) allows the endoplasmic reticulum (ER) to recover from the accumulation of misfolded proteins, in part by increasing its folding capacity. Inositol-requiring enzyme–1 (IRE1) promotes this remodeling by detecting misfolded ER proteins and activating a transcription factor, X-box–binding protein 1, through endonucleolytic cleavage of its messenger RNA (mRNA). Here, we report that IRE1 independently mediates the rapid degradation of a specific subset of mRNAs, based both on their localization to the ER membrane and on the amino acid sequence they encode. This response is well suited to complement other UPR mechanisms because it could selectively halt production of proteins that challenge the ER and clear the translocation and folding machinery for the subsequent remodeling process.

The ER is responsible for the structural maturation of proteins entering the secretory pathway (*1*). When the folding burden on the ER exceeds its capacity, a collection of transcriptional and translational mechanisms termed the UPR is initiated (*2*). The UPR is essential for survival during ER stress and for development in metazoans (*3*, *4*), especially for the differentiation of secretory cells (*5*–*7*). A key sensor of the folding status of the ER is IRE1, a conserved transmembrane protein with an ER luminal sensor domain and cytosolic kinase and ribonuclease domains (*8*, *9*). Accumulation of misfolded proteins in the ER leads to the autophosphorylation and activation of IRE1, which cleaves specific sites in the mRNA encoding the transcription factor X-box–binding protein 1 (XBP-1) (*10*, *11*). This cleavage initiates an unconventional splicing reaction, leading to production of an active transcription factor. Although the sole output of IRE1 activation in yeast appears to be the splicing of the XBP-1 homolog HAC1 (*12*, *13*), several studies in metazoans suggest that IRE1 has additional functions not mediated by XBP-1 (*14*–*16*). Overexpression of IRE1 can also promote cleavage of the 28*S* ribosomal RNA (rRNA) (*17*) and the mRNA encoding IRE1 itself (*18*), which raises the possibility that, under some circumstances, the IRE1 nuclease can act on a broader range of substrates.

To systematically test the XBP-1 dependence of IRE1 outputs, we compared the changes in expression profiles associated with the UPR in *Drosophila* S2 cells depleted of either IRE1 or XBP-1 by RNA interference (RNAi) (Fig. 1, A and B). We induced the UPR with the reducing agent dithiothreitol (DTT), which prevents disulfide-linked folding, and analyzed the changes in expression of ~5000 genes using spotted DNA micro-

Department of Cellular and Molecular Pharmacology, University of California San Francisco, Howard Hughes Medical Institute, and the California Institute for Quantitative Biomedical Research, University of California San Francisco, San Francisco, CA 94143, USA.

*To whom correspondence should be addressed. E-mail: weissman@cmp.ucsf.edu

arrays (*19*). Among strongly induced genes, there was excellent overlap between the requirements for IRE1 and XBP-1 (Fig. 1A). Consistent with previous results (*12*, *15*, *20*, *21*), the IRE1- and XBP-1–dependent induced genes were enriched for ER-related processes, including folding, glycosylation, protein trafficking, and lipid metabolism (table S1).

In addition to this classic induction pathway, our studies revealed an unanticipated branch of the UPR in which IRE1 mediated the rapid and selective repression of ER-targeted mRNAs. Specifically, we observed a cluster of genes whose repression was dependent on IRE1 but not on XBP-1 or the other UPR components ATF6, PERK, and ATF4 (Fig. 1, B to D and fig. S1). We also observed repression of several of these targets by tunicamycin, which induces the UPR by inhibiting glycosylation (fig. S2). Within the repressed targets, there was a strong enrichment for genes encoding plasma membrane and other secreted proteins (Fig. 1C and table S2), suggesting that these proteins traffic through the ER but are not directly involved in ER function. The repression was fast compared with expression changes mediated by XBP-1, which typically displayed a lag phase of ~2 hours (Fig. 2A). Thus, this repressive response is well-suited to relieve acute ER stress, because it would specifically prevent the translation of proteins targeted to the ER before the protective mechanisms of the XBP-1–dependent pathway could take effect (*22*).

The observed down-regulation appears to be mediated by mRNA degradation rather than transcriptional repression. Blocking transcription with actinomycin D (Fig. 2B and fig. S3) or 5,6-dichloro-1-βD-ribofuranosylbenzimidazole (DRB) (fig. S4) had no effect on the DTT-dependent repression of the target mRNAs. Furthermore, expressing the coding sequence under the control of the copper-inducible metallothionein promoter was sufficient to allow wild-type levels of repression of three target mRNAs (Fig. 2C), ruling out a role for the natural promoters or 5′ and 3′ untranslated regions.

We next investigated whether this mRNA degradation was initiated by internal cleavage, such as one mediated by the IRE1 endonuclease. If this were the case, the resulting RNA fragments would be subject to degradation by housekeeping machinery, including XRN1 and the Ski2-3-8 complex, which are involved in cytoplasmic 5′ to 3′ and 3′ to 5′ degradation, respectively (*23*). Consistent with such a mechanism, depletion of XRN1 by RNAi led to the IRE1- and DTT-dependent accumulation of 3′ fragments of the target mRNAs SPARC (Fig. 3, A and B) and TMS1 (fig. S5). Similarly, depletion of Ski2 led to the IRE1- and DTT-dependent accumulation of 5′ fragments (Fig. 3, A and C, and fig. S5). Primer extension analysis of mRNA derived from XRN1-depleted cells revealed two putative cleavage sites between positions 421 and 422 (codon 104) and 320 and 321 (codon 71) of the SPARC transcript (Fig. 3D). The two sites have similar sequences (Fig. 3E) and are consistent with the sizes of the 5′ SPARC fragments seen in the Ski2 depletion, which suggests that the fragments were generated by two distinct endonucleolytic cleavage events.

Exploration of the cis elements necessary for targeting of mRNAs to this decay pathway revealed a critical role for an ER-targeting signal sequence. Deletion of the region encoding residues 1 to 40 in a reporter expressing SPARC abolished DTT-dependent degradation (Fig. 4A). This defect was due to the removal of the N-terminal signal sequence, because replacement of these residues with the signal sequence of BiP, an ER chaperone that is not subject to IRE1-mediated degradation, completely restored the regulation (Fig. 4A). Replacement of three hydrophobic residues in the SPARC signal sequence with charged residues also abolished DTT regulation, whereas proximal silent mutations had no effect (Fig. 4B and

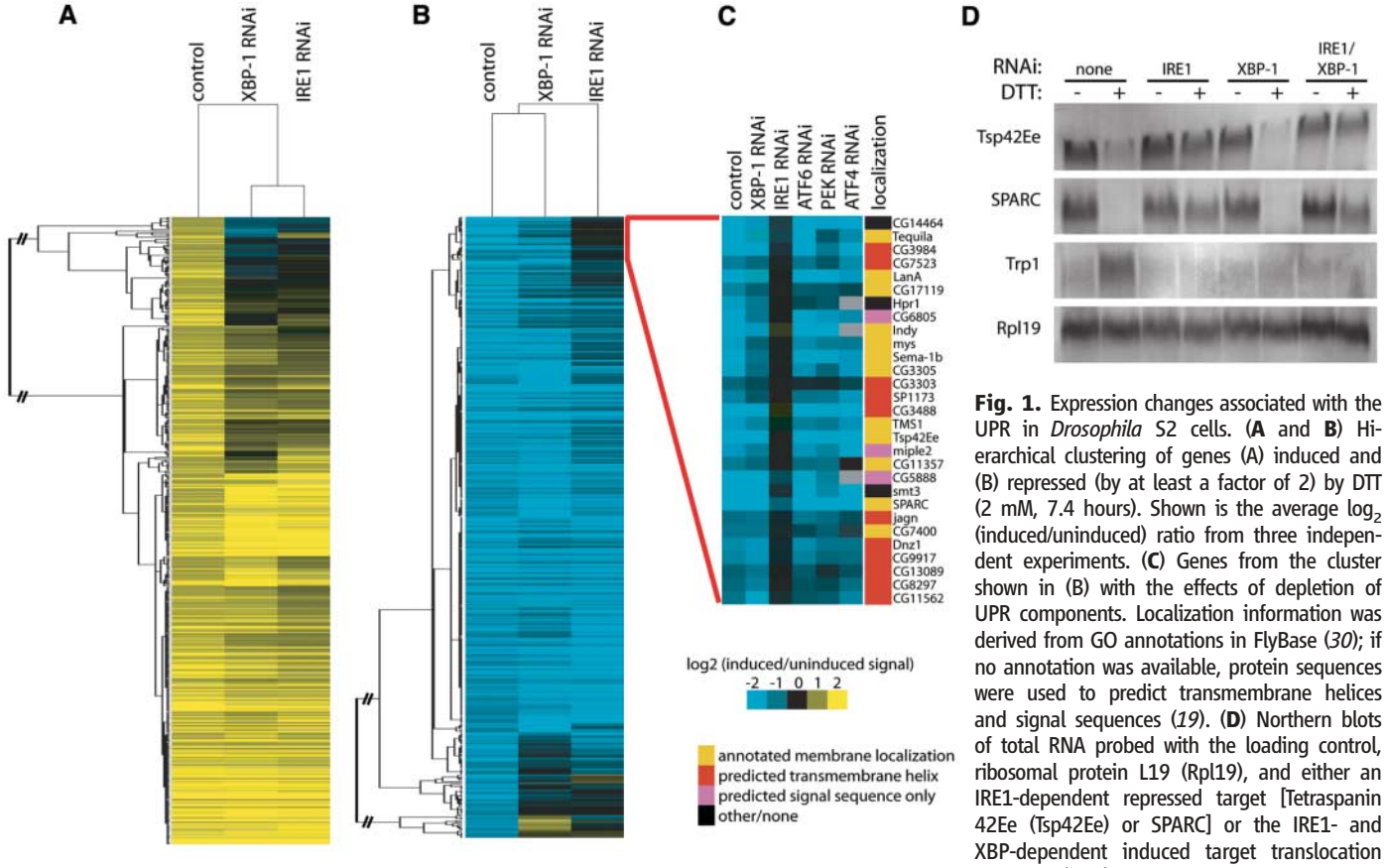


Fig. 1. Expression changes associated with the UPR in *Drosophila* S2 cells. (**A** and **B**) Hierarchical clustering of genes (A) induced and (B) repressed (by at least a factor of 2) by DTT (2 mM, 7.4 hours). Shown is the average $\log_2$ (induced/uninduced) ratio from three independent experiments. (**C**) Genes from the cluster shown in (B) with the effects of depletion of UPR components. Localization information was derived from GO annotations in FlyBase (*30*); if no annotation was available, protein sequences were used to predict transmembrane helices and signal sequences (*19*). (**D**) Northern blots of total RNA probed with the loading control, ribosomal protein L19 (Rpl19), and either an IRE1-dependent repressed target [Tetraspanin 42Ee (Tsp42Ee) or SPARC] or the IRE1- and XBP-dependent induced target translocation protein 1 (Trp1).

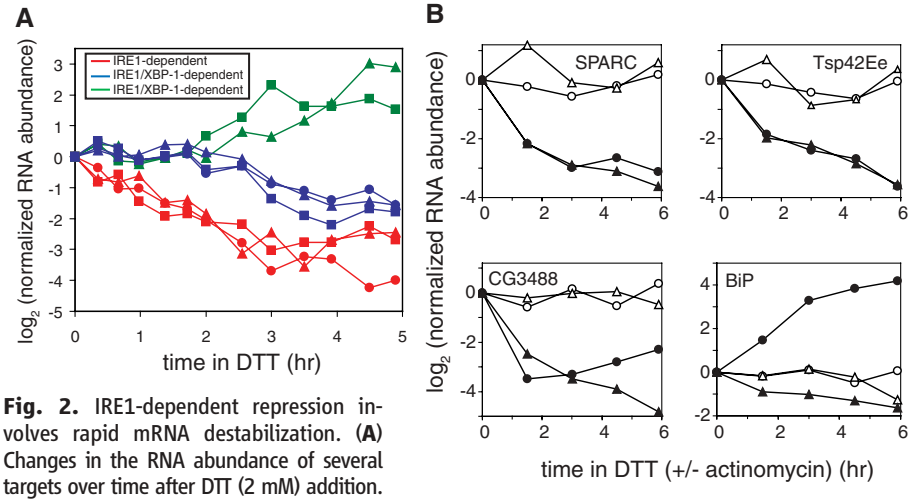


Fig. 2. IRE1-dependent repression involves rapid mRNA destabilization. (**A**) Changes in the RNA abundance of several targets over time after DTT (2 mM) addition. Red circles, SPARC; red triangles, Tsp42Ee; red squares, CG3488; blue circles, Akap200; blue triangles, Stat92E; blue squares, dacapo; green triangles, Trp1; and green squares, CG8286. Results are representative of time-course data measured by microarray. (**B**) Changes in RNA abundance over time after DTT addition, in the presence or absence of the transcription inhibitor actinomycin D. Cells received no treatment (open circles), actinomycin D (open triangles), DTT (black circles), or both (black triangles). (**C**) Repression of endogenous transcripts (white bars) and transcripts of transfected reporters (gray bars) containing the coding sequences of the corresponding targets under the control of the metallothionein (MT) promoter, after 4.5 hours of DTT treatment. The averages and SDs of two to four independent transfections are shown. Measurements for (A to C) were done by quantitative reverse transcription polymerase chain reaction (RT-PCR).

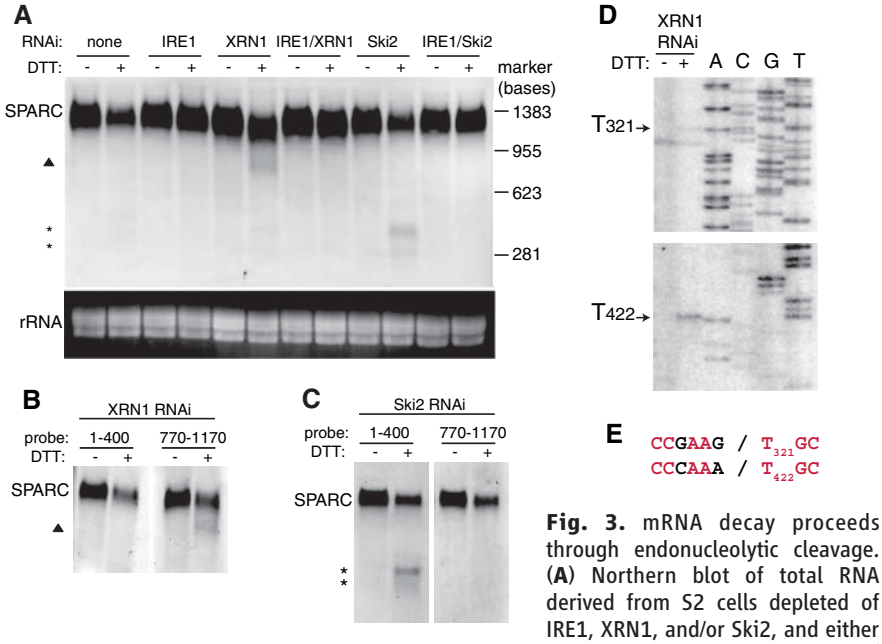


Fig. 3. mRNA decay proceeds through endonucleolytic cleavage. (**A**) Northern blot of total RNA derived from S2 cells depleted of IRE1, XRN1, and/or Ski2, and either untreated or treated with DTT (2 mM, 1.5 hours), hybridized with a probe complementary to the 5′ UTR and coding sequence of SPARC. The triangle indicates the 3′ mRNA fragment, and asterisks indicate the 5′ fragments. (**B** and **C**) Northern blots as in (A) hybridized with probes complementary to the 5′ or 3′ end of the SPARC transcript. Numbers refer to the nucleotide position within the transcript (total length, 1170 nucleotides). (**D**) Primer extension analysis of poly(A)$^+$ RNA isolated from cells depleted of XRN1 and either untreated or treated with DTT as in (A and B), using a radiolabeled primer for SPARC. The two DTT-specific bands (positions T_{321} and T_{422} in the transcript) were seen in two independent experiments. (**E**) Sequences surrounding the sites identified in (D).

fig. S6). Examination of the effects of these mutations on localization of the protein, using C-terminal green fluorescent protein (GFP) fusions to the first 138 amino acids of SPARC [SPARC(Δ139–304)], confirmed that the charged, but not the silent, mutations disrupted signal sequence function (Fig. 4C). Similar results were seen with a second target, CG9917 (Fig. 4B), which suggested that localization of the mRNA and nascent chain to the ER membrane is an important aspect of the degradation process. Apart from the signal sequence, it appears that cis elements within the SPARC mRNA are distributed and redundant, such that no discrete loss of regulation was seen on removal of individual RNA sequence elements (Fig. 4A and SOM Text).

The requirement for a functional signal sequence suggests that IRE1-mediated mRNA decay occurs during cotranslational translocation. This led us to ask whether the nascent protein was involved in transcript degradation. To address this possibility, we took advantage of the fact that for the SPARC(Δ139–304) reporter, insertion of one nucleotide after codon 63 or two nucleotides after codon 88 gave rise to frame-shifted messages that fortuitously lacked downstream stop codons. We fused these to the GFP coding sequence in the appropriate frame. Despite the fact that these constructs expressed ER-targeted proteins (Fig. 4E), the presence of the frame-shift mutations rendered the transcripts largely resistant to IRE1-dependent degradation (Fig. 4D and fig. S7). This effect was unlikely to be caused by the disruption of RNA sequence or structure, because constructs with three nucleotide insertions retained their regulation (Fig. 4D). Thus, it appears that, in addition to being necessary for targeting to the ER membrane, correct translation of the polypeptide plays an important role in the regulation of this mRNA.

The features of the IRE1-mediated mRNA decay pathway described here are interesting in light of the known properties of IRE1. Signal sequence–mediated targeting of the messages would bring them to the cytosolic face of the ER membrane, perhaps in proximity to IRE1's nuclease domain. Because degradation appears to be initiated by an endonucleolytic cleavage, IRE1 may act directly on the target messages, showing a reduced specificity similar to its closest homolog, ribonuclease (RNase) L (*24*). Alternatively, IRE1 may rapidly recruit or activate a second ribonuclease or may promote translational stalling and cleavage by no-go decay (*25*). More speculatively, the requirement for in-frame translation in the degradation of these mRNAs suggests that there is an active discrimination among translocating polypeptides, based on their physical properties or

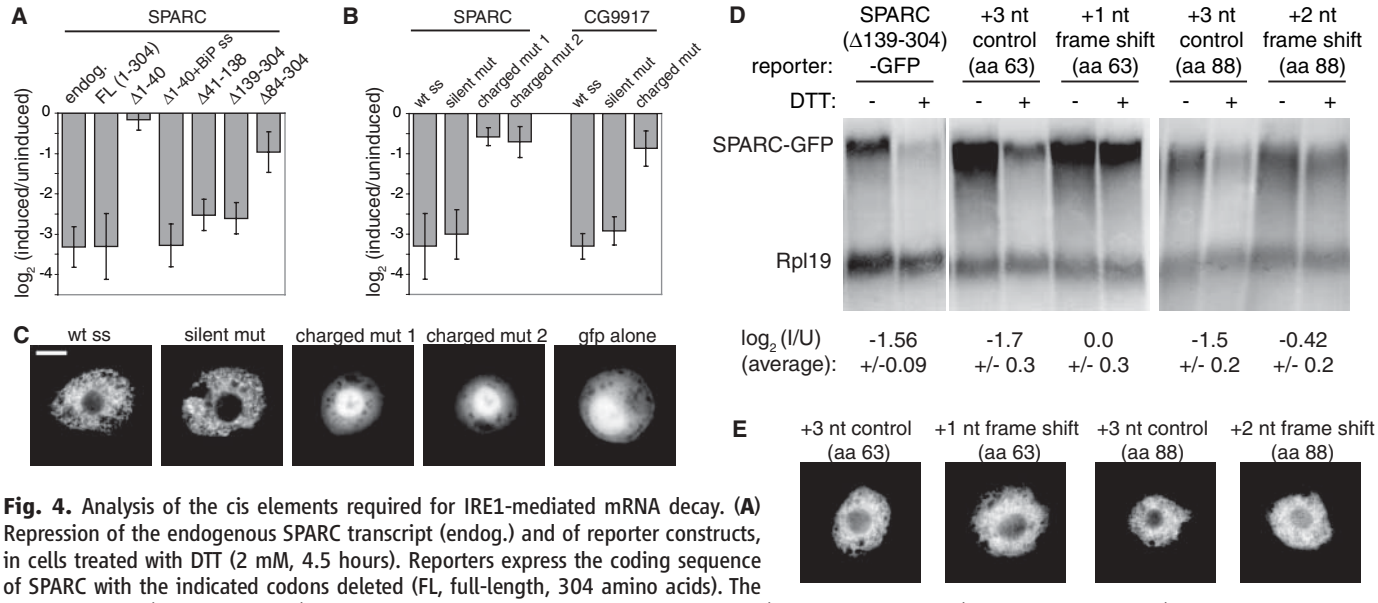


Fig. 4. Analysis of the cis elements required for IRE1-mediated mRNA decay. (**A**) Repression of the endogenous SPARC transcript (endog.) and of reporter constructs, in cells treated with DTT (2 mM, 4.5 hours). Reporters express the coding sequence of SPARC with the indicated codons deleted (FL, full-length, 304 amino acids). The Δ1–40+BiP ss (signal sequence) construct contains the coding sequences of BiP (amino acids 1 to 23) followed by SPARC (amino acids 41 to 304). Reporter repression in (A and B) was measured by real-time quantitative PCR; shown are the averages and SDs of three to five independent transfections. (**B**) Repression of reporter constructs in cells treated as in (A). Reporters express the FL coding sequence of either SPARC or CG9917, with either the wild-type signal sequence (wt ss) or with mutated signal sequences. Charged mut constructs replace three or four signal sequence residues with charged residues; silent mut constructs contain silent mutations at similar sites (see fig. S6). (**C**) Images of live S2 cells expressing SPARC(Δ139–304)–GFP with the signal sequence mutations in (B). All reporters were regulated similarly to their full-length, untagged counterparts. Scale bar, 10 μm. (**D**) Northern blots of SPARC(Δ139–304)-GFP with and without insertions of one to three nucleotides after the indicated amino acid. Numbers below refer to the average (± SD) repression of these constructs in three to five independent transfections, measured by quantitative RT-PCR. (**E**) Images of cells expressing reporters from (D). Scale is the same as in (C).

propensity for translational stalling. For example, binding of translocating peptides to activated IRE1 via its luminal hydrophobic cleft (*26*) could mediate the recruitment of translating ribosomes, which would in turn expose the mRNAs to the nuclease domain of IRE1. A targeting mechanism based on direct peptide recognition or translational effects would potentially allow IRE1 to focus on messages that present the most immediate challenge to the translocation and folding machinery.

How might the mRNA degradation pathway described here complement the previously characterized transcriptional and translational branches of the UPR in allowing for a coherent response to misfolded proteins? Under stress conditions, the burden on the ER exceeds its capacity to fold proteins, yet the solution, to synthesize more ER folding machinery, will temporarily add to that burden. Our studies reveal that IRE1 initiates two distinct cascades. The rapid, XBP-1–independent pathway, together with the generalized translational inhibition mediated by PERK (*27*), can immediately relieve the burden on the ER and partially clear the translocation machinery, freeing up resources for the subsequent up-regulation of genes directly involved in folding and trafficking by the slower, IRE1- and XBP-1–dependent pathway. More broadly, given the increasing evidence for large-scale intracellular localization of mRNAs (*28*), the coupling of RNA degradation to local stimuli may prove to be an efficient and widely used strategy for rapidly altering the local distribution of proteins.

References and Notes

1. L. Ellgaard, A. Helenius, *Nat. Rev. Mol. Cell Biol.* **4**, 181 (2003).
2. M. Schroder, R. J. Kaufman, *Annu. Rev. Biochem.* **74**, 739 (2005).
3. A. M. Reimold *et al.*, *Genes Dev.* **14**, 152 (2000).
4. X. Shen *et al.*, *Cell* **107**, 893 (2001).
5. H. P. Harding *et al.*, *Mol. Cell* **7**, 1153 (2001).
6. P. Zhang *et al.*, *Mol. Cell. Biol.* **22**, 3864 (2002).
7. N. N. Iwakoshi *et al.*, *Nat. Immunol.* **4**, 321 (2003).
8. J. S. Cox, C. E. Shamu, P. Walter, *Cell* **73**, 1197 (1993).
9. K. Mori, W. Ma, M. J. Gething, J. Sambrook, *Cell* **74**, 743 (1993).
10. M. Calfon *et al.*, *Nature* **415**, 92 (2002).
11. H. Yoshida, T. Matsui, A. Yamamoto, T. Okada, K. Mori, *Cell* **107**, 881 (2001).
12. K. J. Travers *et al.*, *Cell* **101**, 249 (2000).
13. M. Niwa, C. K. Patil, J. DeRisi, P. Walter, *Genome Biol.* **6**, R3 (2005).
14. K. Zhang *et al.*, *J. Clin. Invest.* **115**, 268 (2005).
15. X. Shen, R. E. Ellis, K. Sakaki, R. J. Kaufman, *PLoS Genet.* **1**, e37 (2005).
16. U. Ozcan *et al.*, *Science* **306**, 457 (2004).
17. T. Iwawaki *et al.*, *Nat. Cell Biol.* **3**, 158 (2001).
18. W. Tirasophon, K. Lee, B. Callaghan, A. Welihinda, R. J. Kaufman, *Genes Dev.* **14**, 2725 (2000).
19. Materials and methods are available as supporting material on *Science* Online.
20. A. H. Lee, N. N. Iwakoshi, L. H. Glimcher, *Mol. Cell. Biol.* **23**, 7448 (2003).
21. F. Urano *et al.*, *J. Cell Biol.* **158**, 639 (2002).
22. In yeast, down-regulation of mRNAs encoding membrane proteins is mediated by the classic Ire1p/Hac1p pathway (*29*), which may reflect a rapid response via transcriptional repression allowed by the relatively short half-lives of messages in yeast.
23. D. Gatfield, E. Izaurralde, *Nature* **429**, 575 (2004).
24. B. Dong, M. Niwa, P. Walter, R. H. Silverman, *RNA* **7**, 361 (2001).
25. M. K. Doma, R. Parker, *Nature* **440**, 561 (2006).
26. J. J. Credle, J. S. Finer-Moore, F. R. Papa, R. M. Stroud, P. Walter, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18773 (2005).
27. H. P. Harding, Y. Zhang, A. Bertolotti, H. Zeng, D. Ron, *Mol. Cell* **5**, 897 (2000).
28. A. P. Gerber, D. Herschlag, P. O. Brown, *PLoS Biol.* **2**, E79 (2004).
29. Y. Kimata, Y. Ishiwata-Kimata, S. Yamada, K. Kohno, *Genes Cells* **11**, 59 (2006).
30. R. A. Drysdale, M. A. Crosby, *Nucleic Acids Res.* **33**, D390 (2005).

31. We thank G. Goshima, A. Carroll, N. Stuurman, and K. Shepard for assistance in making the microarrays; S. Rogers, T. Kress, R. Kelso, and J. DeRisi and members of his laboratory for technical advice; and P. Walter, F. Papa, and members of the laboratories of P.W. and J.S.W. for discussions. This work was supported by funds from the Sandler family, the Howard Hughes Medical Institute (J.S.W.) and a Ruth L. Kirschstein National Research Service Award (J.H.). Microarray data are available at the Gene Expression Omnibus (GEO), accession no. GSE4867.

Supporting Online Material

www.sciencemag.org/cgi/content/full/313/5783/104/DC1
Materials and Methods
SOM Text
Figs. S1 to S7
Tables S1 and S2
References

8 May 2006; accepted 24 May 2006
10.1126/science.1129631

Polo-Like Kinase Cdc5 Controls the Local Activation of Rho1 to Promote Cytokinesis

Satoshi Yoshida,[1] Keiko Kono,[1,2] Drew M. Lowery,[3] Sara Bartolini,[1] Michael B. Yaffe,[3] Yoshikazu Ohya,[2] David Pellman[1]*

The links between the cell cycle machinery and the cytoskeletal proteins controlling cytokinesis are poorly understood. The small guanine nucleotide triphosphate (GTP)–binding protein RhoA stimulates type II myosin contractility and formin-dependent assembly of the cytokinetic actin contractile ring. We found that budding yeast Polo-like kinase Cdc5 controls the targeting and activation of Rho1 (RhoA) at the division site via Rho1 guanine nucleotide exchange factors. This role of Cdc5 (Polo-like kinase) in regulating Rho1 is likely to be relevant to cytokinesis and asymmetric cell division in other organisms.

Cytokinesis, the physical separation of daughter cells after mitosis, involves the dynamic reorganization of the cortical cytoskeleton and is precisely regulated both in time and space (*1*, *2*). Successful cytokinesis is critical to the maintenance of genome stability: Cytokinesis failure can lead to chromosome aberrations and cancer (*3*). The timing and mechanism of contractile actin ring (CAR) assembly is an important but poorly understood aspect of the eukaryotic cell cycle.

The Polo-like kinase Cdc5 controls many aspects of cell division, including cytokinesis (*4*, *5*). Polo-like kinase is involved in cytokinesis in animal cells, but its precise role has been difficult to address because manipulations that inhibit Polo-like kinase produce spindle assembly defects, which affect CAR organization (*2*, *6*, *7*). RhoA is an attractive candidate to be under Polo-like kinase regulation because of its essential role in CAR assembly and contraction (*1*, *2*, *8*, *9*). In yeast, cytokinesis is largely independent of microtubules (*8*), making yeast an advantageous system to study the role of Polo-like kinase. Additionally, the contractile ring is not essential in most budding yeast strains; another mechanism involving cell wall deposition can substitute for CAR function (*10*, *11*). This facilitates genetic analysis because null alleles affecting CAR function can be studied in viable cells. Despite these differences, the core machinery for CAR assembly is conserved in budding yeast: CAR assembly requires Rho1, the functional homolog of RhoA in animal cells, which activates actin filament assembly via formins (*8*, *12*).

The timing of CAR assembly was characterized by actin staining and by labeling of the CAR with a fusion between tropomyosin and green fluorescent protein (Tpm2-GFP) (Fig. 1A and fig. S1A). To identify cell cycle regulators required for CAR assembly, we reexamined CAR assembly in mutants that block mitotic exit (*11*, *13*). The CAR assembled in *cdc15-2* or *cdc14-1* strains, mutations that compromise the canonical mitotic exit network (MEN) (Fig. 1A and fig. S1B). By contrast, CAR assembly was impaired in mutants lacking Cdc5 (Fig. 1A and fig. S1B).

The Cdc5 requirement for CAR formation was observed either by using a very severe conditional allele, *cdc5-2*, or by examining the null phenotype after Cdc5 depletion (fig. S2A) (*14*). The defect in CAR assembly in cells lacking Cdc5 was not a secondary consequence

[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute and Division of Hematology/Oncology, Children's Hospital Boston and Harvard Medical School, Boston, MA 02115, USA. [2]Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Chiba 277-8562, Japan. [3]Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

*To whom correspondence should be addressed. E-mail: david_pellman@dfci.harvard.edu

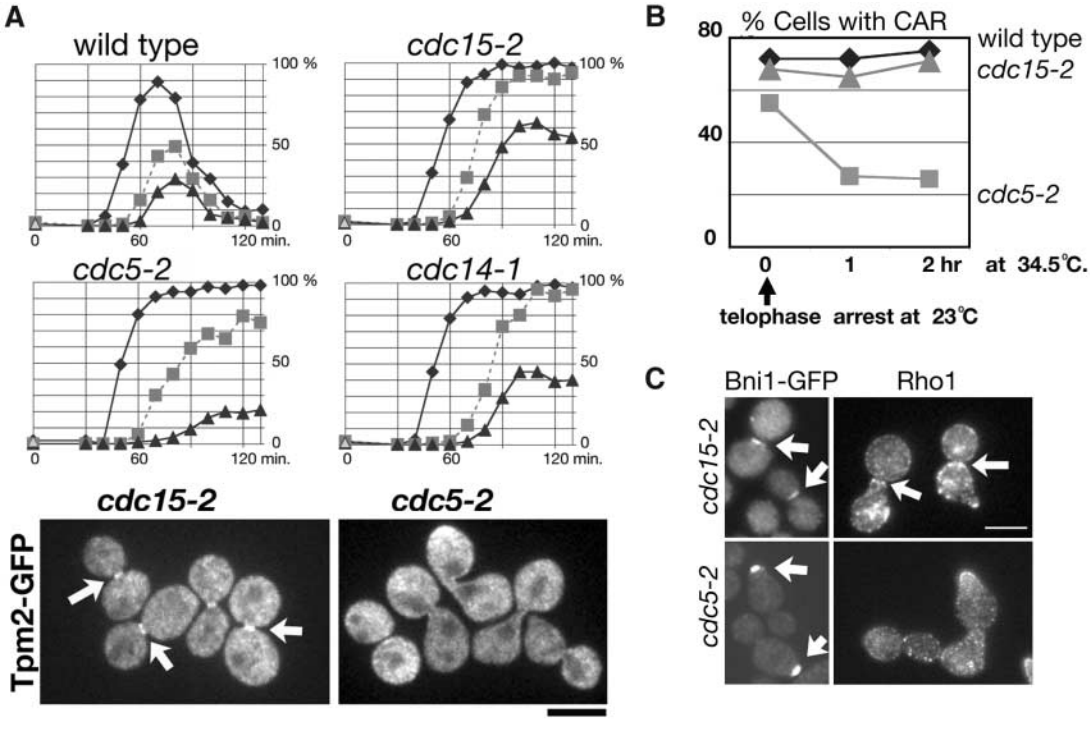


Fig. 1. Cdc5 is required for CAR assembly. (**A**) CAR assembly is defective in *cdc5-* but not *cdc15-* or *cdc14*-arrested cells. Cells were released from a G1 block (α factor) and examined at intervals after release at the nonpermissive temperature 34.5°C for budding index (diamonds), percentage of anaphase cells by labeling nuclear DNA (squares), and bud neck localization of Tpm2-GFP (triangles). The α factor was re-added at 60 min to prevent progression into a second cell cycle. (Bottom) Representative images of cells from the 120-min time point. Scale bar indicates 5 μm. (**B**) Cdc5 is required to maintain the CAR. Log-phase cultures of *cdc5-2* and control cells were arrested in telophase by overexpression of *BFA1* from the *GAL1* promotor for 3 hours. The telophase-arrested cells were shifted to a restrictive temperature (34.5°C) to inactivate *cdc5-2* and *cdc15-2* and fixed, and the CAR formation was visualized with the use of Tpm2-GFP. (**C**) Cdc5 is required for the subcellular targeting of Bni1 and Rho1. Bni1-3xGFP expressed at an endogenous level was visualized after 2-hour incubation of the cells at 34.5°C. Rho1 was visualized by immunofluorescence labeling with a rabbit antibody against Rho1 130 min after release to 34.5°C from a G1 block. Arrows point to the bud neck signals in *cdc15* cells and the bud tip signals in *cdc5* cells.

of the anaphase defect in these cells: Cdc5 was not only required to establish CAR assembly but also to maintain it. When cells were arrested before mitotic exit by overexpression of a MEN inhibitor, Bfa1, CAR maintenance still required Cdc5 [Fig. 1B and Supporting Online Material (SOM) Text]. Furthermore, early anaphase pathways regulated by Cdc5 were not required for CAR assembly (SOM Text and fig. S2B). In the absence of Cdc5, a residual fraction of cells were observed that had thin CARs (Fig. 1A and fig. S2B). Thus, Cdc5, although not absolutely essential for CAR assembly, is an important regulator.

Next we characterized the nature of the CAR assembly defect in *cdc5-2* cells at the restrictive temperature, with the use of the MEN mutant *cdc15-2* as a control. Cdc5 is required for septin disassembly after mitotic exit but is not required for septin assembly (*15*). The septin Shs1, the type II myosin Myo1, and the IQGAP homolog Iqg1 (Cyk1) were properly recruited to the bud neck both in the *cdc5-2* and in the *cdc15-2* control cells (fig. S2C). By contrast, although the formin Bni1 and Rho1 localized normally to the neck in *cdc15-2*–arrested cells, neither Bni1 nor Rho1 was detectable at the bud neck in *cdc5-2*–arrested cells (Fig. 1C). Furthermore, Bni1 was often mistargeted to the bud tip in *cdc5-2*–arrested cells (Fig. 1C). Thus, Cdc5 is required for the neck recruitment of Rho1 and Bni1, a key downstream factor required for CAR assembly.

To identify potential Cdc5 substrates relevant to Rho1 regulation, we screened all known budding yeast Rho guanosine triphosphatase (GTPase) regulators for interaction with the Polo-box domain (PBD) of Cdc5 (*7*, *16*, *17*). PBD binding partners include the Rho1 guanine nucleotide exchange factors (GEFs) Rom2 and Tus1, the Rho1 GTPase activating protein (GAP) Sac7, and a putative Rho-GAP, Ecm25 (Fig. 2, A and B). The PBD also binds Bem3, a GAP for Cdc42, the major small GTPase that controls polarized morphogenesis in budding yeast. These interactions are highly specific to the priming phosphorylation of the candidates, because the interaction was lost with a Polo-box pincer mutation, PBD* (Fig. 2, A and B), that abolishes phosphospecific recognition of the substrates by PBD (*17*). All the candidates had multiple potential Polo-box binding sequences, binding motifs that are often primed by cyclin-dependent kinase (CDK)–dependent phosphorylation (*16*). Tus1, Bem3, Sac7, and Ecm25 are also good CDK substrates (*18*).

Because of the requirement of Rho1 for CAR assembly, we focused our subsequent analysis on the Rho1 GEFs, Tus1 and Rom2 (*19*). Different patterns of synthetic lethal interactions distinguish genes required for CAR function from genes required for the CAR-independent pathway for cytokinesis (*8*, *11*). Analysis of CAR assembly and synthetic lethal interactions strongly implicated Tus1 and Rom2 in cytokinesis, with Tus1 playing the dominant role in CAR assembly (SOM Text, table S1, and fig. S3). These findings are consistent with the requirement for the fission yeast Tus1 homolog in cytokinesis (*20*–*22*). The pattern of synthetic lethality observed with *cdc5-2* also provided genetic evidence for a role for Cdc5 in the CAR-dependent mechanism (SOM Text and fig. S4).

We next evaluated the possibility that Tus1 is a direct Polo-like kinase substrate. The N terminus of Tus1 ($Tus1^{1\text{-}300}$), containing two Polo-box binding motifs (Ser-Ser/Thr-Pro), is sufficient for Polo-box binding (fig. S5, A and B). The serine residues immediately preceding the CDK consensus site in the Polo-box binding motifs of Tus1 (Ser^{7} and Ser^{92}) were mutated to threonine (hereafter referred to as ST mutations), a well-characterized change that disrupts binding to the PBD (*16*, *17*) but that should not affect CDK-dependent phosphorylation. Glutathione *S*-transferase (GST)–PBD precipitation experiments revealed that neither $Tus1\text{-}ST^{1\text{–}300}$ nor full-length Tus1-ST protein bound to the PBD (Fig. 2C and fig. S5B). Similar ST substitutions were introduced into the coding sequence for Rom2; the Rom2-ST protein also had a reduced affinity for the Cdc5 PBD (fig. S6). Thus, Tus1 and Rom2 interact with Cdc5 by a mechanism similar to that of other known Polo substrates.

Biochemical experiments suggested that Tus1 is a bona fide in vivo Cdc5 substrate. Although we were not able to detect a clear mobility shift of the full-length Tus1 (1307 amino acids), the Tus1 Polo-box interacting

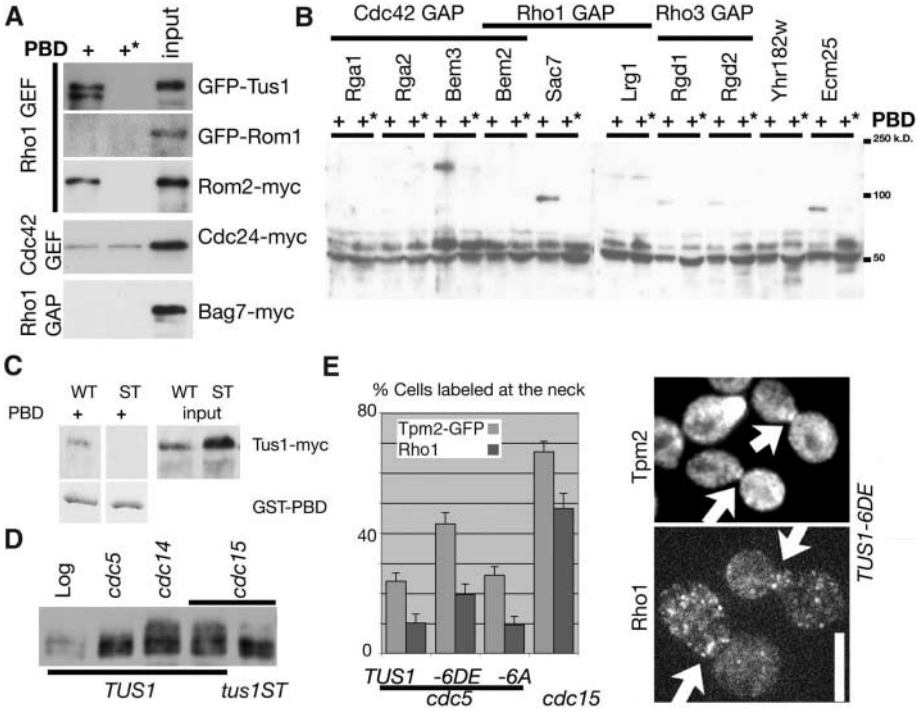


Fig. 2. A screen for Polo-box binding partners suggests that Cdc5 controls multiple Rho-type GTPase regulators. All potential Rho GTPase regulators in the yeast genome were screened for physical interaction with the Cdc5 PBD and a control pincer mutation (PBD*). (**A**) In cases where epitope-tagged constructs were not available from the yeast tandem affinity purification (TAP)–tagged library (*24*), epitope-tagged constructs were generated. Tus1 and Rom1 were overexpressed from the *GAL1* promoter for this initial screen. After precipitation assay, Western blotting was performed to detect the indicated epitope tags of the bound proteins. + indicates the wild-type GST-PBD; +* indicates the pincer mutant. (**B**) PBD or PBD* precipitation assays with lysates from strains from the yeast TAP-tag fusion library. (**C**) Endogenous-level Tus1 binds to the PBD. This binding is abolished by the ST mutation and thus requires the Polo-box binding motif. WT, wild type. (**D**) Hyperphosphorylation of $Tus1^{1\text{-}300}$-myc requires Cdc5 and the Polo-box binding motifs. Shown are Western blots to detect endogenous-level $Tus1^{1\text{-}300}$-myc or its ST derivative in the indicated strains after 2.5 hours at the restrictive temperature of 34.5°C. (**E**) Phosphomimetic mutations at Polo-consensus phosphorylation sites partially bypass the requirement for Cdc5 for CAR assembly and Rho1 recruitment. The indicated strains were arrested with α factor and released at the nonpermissive temperature (33.5°C) for 140 min. The CAR was visualized with the use of Tpm2-GFP, and Rho1 was visualized by immunofluorescence. More than 400 cells were counted for each sample. The experiment was repeated three times with near-identical results. By a Student's *t* test, the difference in Tpm2-GFP labeling in cells expressing Tus1-6DE versus Tus1 was significant ($P = 0.011$). Error bars indicate SEM. Scale bar, 5 μm; arrows point to the bud neck signal.

domain (PID), Tus1$^{1\text{-}300}$, was phosphorylated in vivo as evidenced by a motility shift that was abolished by phosphatase treatment in *cdc15-2*–arrested cells (figs. S2D and S5C). Most notably, the slowest migrating form of Tus1$^{1\text{-}300}$ (hereafter referred to as hyperphosphorylated) was absent in *cdc5-2*–arrested cells (Fig. 2D). This hyperphosphorylated form could also be induced in cycling cells by overexpression of Cdc5 (fig. S5D). Furthermore, Tus1-ST$^{1\text{-}300}$ also lost this hyperphosphorylated band (Fig. 2D). Thus, Tus1$^{1\text{-}300}$ is phosphorylated in late mitosis in a manner that depends on Cdc5. The fact that this Cdc5-dependent phosphorylation also required physical interaction between Tus1$^{1\text{-}300}$ and the Polo box strongly suggests that the Cdc5-dependent phosphorylation of Tus1$^{1\text{-}300}$ is direct. Indeed, mass spectrometry confirmed that the Tus1$^{1\text{-}300}$ was readily phosphorylated by a human Polo-like kinase, Plk1, in vitro (table S2), identifying at least 12 phosphorylated residues. Although we were not able to map phosphorylation sites directly on full-length Tus1 because of its low expression (*23*), together with our genetic data these findings strongly suggest that Tus1 is a Cdc5 substrate.

We attempted similar experiments to characterize the potential Polo-like kinase dependent phosphorylation of Rom2. We identified several in vitro Plk1 phosphorylation sites by mass spectrometry (table S3). However, we were not able to identify a mobility shift for full-length Rom2, nor were we able to detect a mobility shift for several N-terminal Rom2 fragments. These fragments were not well expressed in *Escherichia coli* and thus may not fold properly. Nevertheless, the similar phenotypes of *rom2-ST* and *tus1-ST* strains (SOM Text, table S4, and fig. S7) suggest that in vivo Cdc5 regulates Rom2 similarly to Tus1.

A Tus1 protein containing several phosphomimetic mutations at putative Cdc5 phosphorylation sites provided further evidence for the importance of Cdc5-dependent phosphorylation of Tus1 for CAR assembly. Phosphomimetic mutations (Ser to Asp or Thr to Glu) were introduced at Ser or Thr residues within Plk1 phosphorylation consensus sites (Asp/Glu-X-Ser/Thr, where X is any amino acid). This consensus is optimal for Plk1-dependent phosphorylation of vertebrate Cdc25 (*23*) and perfectly matches the in vivo Cdc5 target phosphorylation sites on Mcd1 (Scc1) (*14*). Three out of four Plk1 phosphorylation consensus sites within Tus1$^{1\text{-}300}$ were phosphorylated by Plk1 in vitro (table S2). A Tus1 mutant containing six putative phosphomimetic changes (Tus1-6DE) partially bypassed the requirement for Cdc5, both for CAR formation and for Rho1 localization to the division site (Fig. 2E). A control Tus1 mutant where the putative phosphorylation sites were mutated to Ala failed to restore CAR assembly and Rho1 localization in *cdc5-2* cells, demonstrating that only the negatively charged residues caused a gain-of-function effect at these sites. The magnitude of the gain-of-function phenotype induced by *TUS1-6DE* is likely to be an underestimate, because the introduction of negatively charged residues appears to reduce the already low steady state expression of Tus1 (fig. S5E) (*24*).

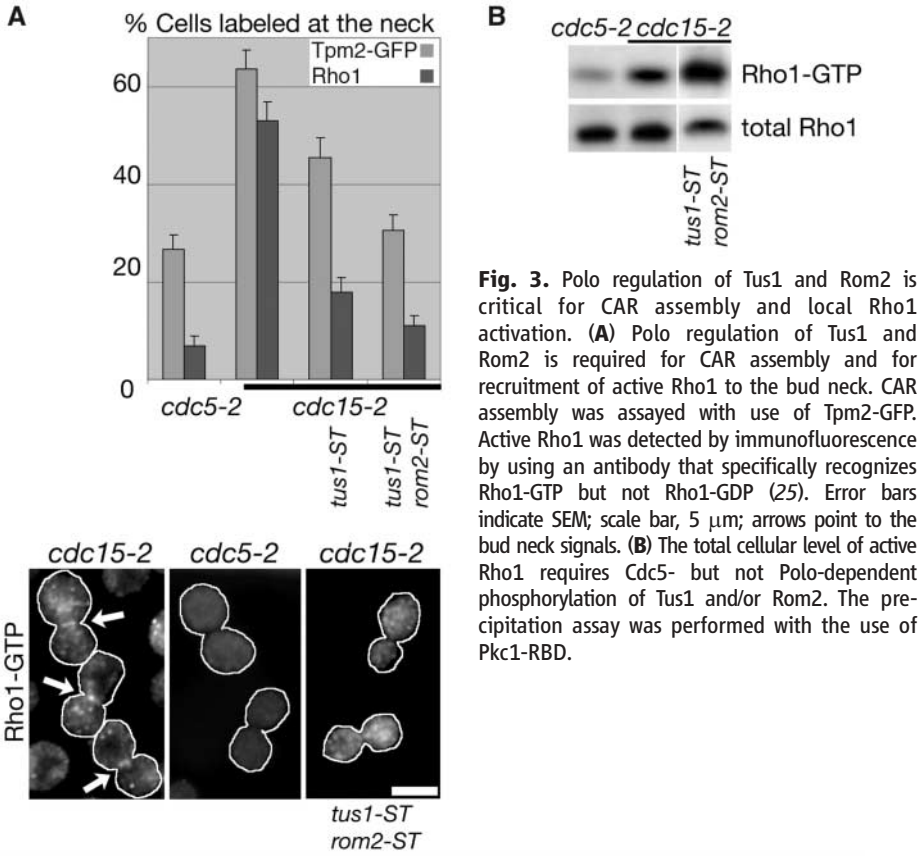


Fig. 3. Polo regulation of Tus1 and Rom2 is critical for CAR assembly and local Rho1 activation. (**A**) Polo regulation of Tus1 and Rom2 is required for CAR assembly and for recruitment of active Rho1 to the bud neck. CAR assembly was assayed with use of Tpm2-GFP. Active Rho1 was detected by immunofluorescence by using an antibody that specifically recognizes Rho1-GTP but not Rho1-GDP (*25*). Error bars indicate SEM; scale bar, 5 μm; arrows point to the bud neck signals. (**B**) The total cellular level of active Rho1 requires Cdc5- but not Polo-dependent phosphorylation of Tus1 and/or Rom2. The precipitation assay was performed with the use of Pkc1-RBD.

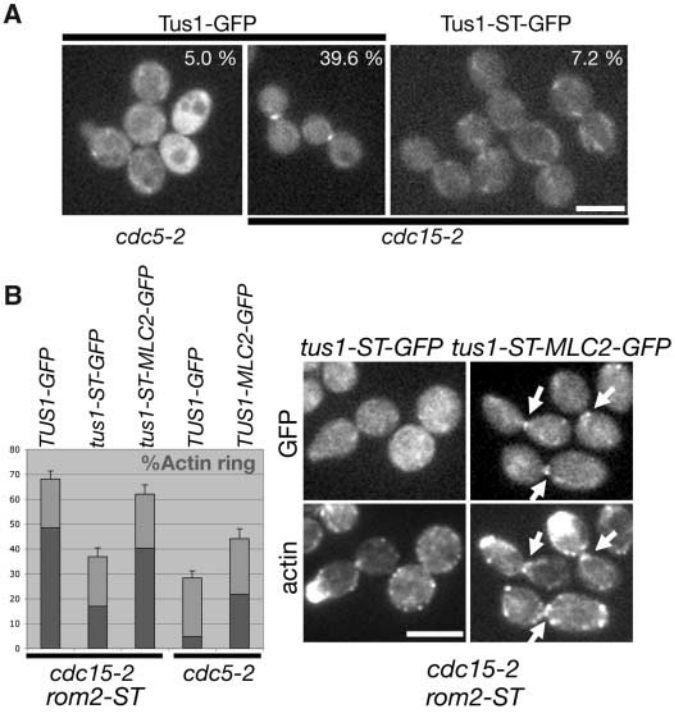


Fig. 4. Cdc5 targets Tus1 to the division site. (**A**) Tus1-GFP requires Cdc5 and its Polo-box binding motifs to localize to the bud neck. Tus1-GFP or the ST derivatives were imaged in *cdc5-2* or in *cdc15-2* 2 hours after incubation at 34.5°C (n > 100), along with the percentages of the large budded cells with GFP signal at the bud neck. Scale bar, 5 μm. (**B**) Tethering Tus1-ST to the bud neck by fusion to Mlc2 corrects the CAR assembly defect of *tus1-ST rom2-ST* strains and partially bypasses the requirement for Cdc5 in CAR assembly. (Left) Numerical summary of the data. The bars indicate the percentage of cells with detectable contractile rings. Error bars indicate SEM. The dark bars indicate the percentage of cells with robust contractile rings. (Right) Examples of cells. (Top) GFP imaging of the indicated fusions. (Bottom) Actin was visualized with alexa568-phalloidin. Arrows indicate cells with robust CAR labeling.

The synthetic lethal interactions observed with *tus1-ST* and *rom2-ST* suggested that these mutations were specifically defective in the CAR-dependent pathway for cytokinesis but not other Rho1-dependent cellular functions (SOM Text, table S4, and fig. S7). Thus, the ST mutants are separation-of-function alleles that are specifically defective in contractile ring function. Supporting this hypothesis, *tus1-ST* strains arrested in telophase were defective in CAR assembly, and this defect was enhanced in *tus1-ST rom2-ST* double mutant strains (Fig. 3A).

Our experiments suggest that Cdc5 controls Rho1 through Tus1 and Rom2. This was directly tested by monitoring the amount of active Rho1 both globally and locally at the bud neck. The total amount of active GTP-bound Rho1 was detected by a precipitation assay with the Rho1-binding domain (RBD) of Pkc1 (Fig. 3B). The amount of active Rho1 was diminished in *cdc5-2*–arrested cells in comparison with *cdc15-2*–arrested cells (Fig. 3B). We also visualized active Rho1 by immunofluorescence using an antibody that specifically detects GTP-bound Rho1 (*25*). Consistent with the precipitation experiment of active Rho1, activated Rho1 was barely detectable in *cdc5-2*–arrested cells but was abundant and recruited to the bud neck region in *cdc15-2*–arrested cells (Fig. 3A). By contrast, *tus1-ST rom2-ST* cells arrested at the *cdc15-2* block had active Rho1 at the bud cortex but lacked active Rho1 at the bud neck (Fig. 3A). Consistent with this observation, the precipitation assay indicated that Rho1 activity was not diminished in *tus1-ST rom2-ST* strains and in fact was slightly elevated (Fig. 3B). The fact that total Rho1 activity was high in *tus1-ST rom2-ST* strains but low in *cdc5-2* strains suggests that Cdc5 also controls global Rho1 activity through another mechanism, perhaps through Rho1 GAPs such as Sac7 (Fig. 2B). Thus, Cdc5-dependent regulation of Tus1 and Rom2 controls the local activation of Rho1 at the division site but not the global levels of active Rho1 in the cell.

Next, we determined whether Cdc5-dependent phosphorylation is required to recruit Tus1 and Rom2 to the bud neck. Consistent with its important role in CAR assembly, Tus1-GFP was recruited to the bud neck just before CAR constriction and formed a single ring that colocalized and contracted with Myo1-CFP (fig. S8). The bud neck localization of Tus1 was dependent on Cdc5 activity but not on other MEN components (Fig. 4A). Furthermore, Tus1-ST-GFP failed to localize to the division site in *cdc15-2*–arrested cells that had high Cdc5 activity (Fig. 4A).

We next tested whether the requirement for Cdc5 regulation of Tus1 could be bypassed by tethering the ST variant to the bud neck. The coding sequence for Tus1-ST-GFP was fused to the coding sequence for Mlc2, a nonessential type II myosin light chain (*26*), and expressed from the endogenous *TUS1* promoter. Indeed, in *cdc15-2*–arrested cells, Tus1-ST-Mlc2-GFP protein localized to the bud neck and activated CAR assembly, unlike Tus1-ST-GFP (Fig. 4B). Furthermore, Tus1-Mlc2-GFP could partially restore CAR assembly in *cdc5-2*–arrested cells (Fig. 4B). Thus, a critical function of the Cdc5-dependent regulation of cytokinesis is to recruit Tus1, and thus Rho1, to the site of CAR assembly. These results do not exclude the additional possibility that Cdc5 also affects Tus1 GEF activity.

Our study reveals a molecular pathway by which Cdc5 controls contractile ring assembly in budding yeast. Cdc5 is required for the recruitment of Rho GEFs to the division site that in turn is necessary for recruitment and activation of Rho1. The failure of this mechanism results in the mistargeting of the formin Bni1 from the bud neck to the bud tip. This causes a defect in CAR assembly, because the actin filaments in the contractile ring are assembled by formins (*1*, *2*, *8*). Because of the conservation of many components in the pathway and because the RhoA GEF Ect2 is a substrate for Plk1 in vitro (*27*), the recruitment of Rho1 activity to the division site by Polo-like kinases might be conserved in animal cells.

It is possible that Cdc5 has more broad roles in Rho-type GTPase regulation than the local activation of Rho1 at the division site. Cdc5 is required for the global activation of Rho1 in late mitosis, whereas the Cdc5 regulation of Tus1 and Rom2 is required for local Rho1 recruitment and activation. Cdc5 could exert additional control of Rho1 via Rho GAPs. It is also possible that Cdc5 could exert Polo-box independent regulation of Rho1 GEFs. Lastly, given that fission yeast Polo kinase was recently shown to be required for stress-induced polarized growth (*28*), it is intriguing that our biochemical screen identified Bem3, a Cdc42 GAP, as a Polo-box binding partner. Our findings thus elucidate a mechanism by which Polo-like kinase governs Rho1 during cytokinesis and raise the possibility that Polo-like kinase controls other aspects of the cortical cytoskeleton through other Rho GTPases.

References and Notes

1. A. Piekny, M. Werner, M. Glotzer, *Trends Cell Biol.* **15**, 651 (2005).
2. P. P. D'Avino, M. S. Savoian, D. M. Glover, *J. Cell Sci.* **118**, 1549 (2005).
3. T. Fujiwara *et al.*, *Nature* **437**, 1043 (2005).
4. K. S. Lee, J. E. Park, S. Asano, C. J. Park, *Oncogene* **24**, 217 (2005).
5. S. Song, K. S. Lee, *J. Cell Biol.* **152**, 451 (2001).
6. M. A. van Vugt, R. H. Medema, *Oncogene* **24**, 2844 (2005).
7. F. A. Barr, H. H. Sillje, E. A. Nigg, *Nat. Rev. Mol. Cell Biol.* **5**, 429 (2004).
8. M. K. Balasubramanian, E. Bi, M. Glotzer, *Curr. Biol.* **14**, R806 (2004).
9. W. M. Bement, H. A. Benink, G. von Dassow, *J. Cell Biol.* **170**, 91 (2005).
10. M. Schmidt, B. Bowers, A. Varma, D. H. Roh, E. Cabib, *J. Cell Sci.* **115**, 293 (2002).
11. E. A. Vallen, J. Caviston, E. Bi, *Mol. Biol. Cell* **11**, 593 (2000).
12. N. Tolliday, L. VerPlank, R. Li, *Curr. Biol.* **12**, 1864 (2002).
13. J. Lippincott, K. B. Shannon, W. Shou, R. J. Deshaies, R. Li, *J. Cell Sci.* **114**, 1379 (2001).
14. G. Alexandru, F. Uhlmann, K. Mechtler, M. A. Poupart, K. Nasmyth, *Cell* **105**, 459 (2001).
15. C. J. Park *et al.*, *Genetics* **163**, 21 (2003).
16. A. E. H. Elia, L. C. Cantley, M. B. Yaffe, *Science* **299**, 1228 (2003).
17. A. E. Elia *et al.*, *Cell* **115**, 83 (2003).
18. J. A. Ubersax *et al.*, *Nature* **425**, 859 (2003).
19. D. E. Levin, *Microbiol. Mol. Biol. Rev.* **69**, 262 (2005).
20. V. Tajadura, B. Garcia, I. Garcia, P. Garcia, Y. Sanchez, *J. Cell Sci.* **117**, 6163 (2004).
21. T. Mutoh, K. Nakano, I. Mabuchi, *Genes Cells* **10**, 1189 (2005).
22. J. L. Morrell-Falvey, L. Ren, A. Feoktistova, G. D. Haese, K. L. Gould, *J. Cell Sci.* **118**, 5563 (2005).
23. H. Nakajima, F. Toyoshima-Morimoto, E. Taniguchi, E. Nishida, *J. Biol. Chem.* **278**, 25277 (2003).
24. S. Ghaemmaghami *et al.*, *Nature* **425**, 737 (2003).
25. M. Abe, H. Qadota, A. Hirata, Y. Ohya, *J. Cell Biol.* **162**, 85 (2003).
26. J. Luo *et al.*, *J. Cell Biol.* **165**, 843 (2004).
27. F. Niiya, T. Tatsumoto, K. S. Lee, T. Miki, *Oncogene* **25**, 827 (2006).
28. J. Petersen, I. M. Hagan, *Nature* **435**, 507 (2005).

29. We thank M. Abe for generating the antibody specific to GTP-bound Rho1; A. Amon, P. Carvalho, D. Morgan, and I. Sagot for strains and/or reagents; E. Bi, K. Lee, D. Lew, M. Glotzer, I. Mabuchi, A. Toh-e, and members of the Pellman lab for discussions; R. Tomaino and S. Gygi from the Taplin Biological Mass Spectrometry Facility at Harvard Medical School for mass spectrometry; and E. Bi, S. Buttery, S. Godinho, S. Nogami, M. Guillet, and M. Gupta for critically reading the manuscript. S.Y. was supported by a fellowship from Japan Society for Promotion of Science. D.M.L. was supported by a Howard Hughes Medical Institute predoctoral fellowship. M.B.Y. was supported by a NIH grant and a Career Development award from the Burroughs-Wellcome Fund. Y.O. was supported by New Energy and Industrial Technology Development Organization (NEDO) and a grant for Scientific Research from the Ministry of Education, Science, Sports, and Culture of Japan. D.P. was supported by a NIH grant and a Scholar Award from the Leukemia and Lymphoma Society of America. NEDO is a Japanese governmental organization, and there is no conflict of interest.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1126747/DC1
Materials and Methods
SOM Text
Figs. S1 to S8
Tables S1 to S6
References

27 February 2006; accepted 26 May 2006
Published online 8 June 2006;
10.1126/science.1126747
Include this information when citing this paper.